Lösungsmengen von Modellparametrierungen zu nichtnegativen Matrixfaktorisierungsproblemen

Dissertation
zur
Erlangung des akademischen Grades
doctor rerum naturalium
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Rostock

vorgelegt von
Henning Schröder, geb. am 11.11.1987 in Wismar.

23. August 2019

https://doi.org/10.18453/rosdok_id00002546

Gutachter:
Prof. Dr. Klaus Neymeyr, Universität Rostock, Institut für Mathematik
Prof. Dr. Ralf Ludwig, Universität Rostock, Institut für Chemie

Jahr der Einreichung: 2019

Jahr der Verteidigung: 2019

Danksagung

Ich danke den Kollegen der numerischen Mathematik, des Leibniz-Instituts für Katalyse und der Evonik, mit denen sowohl aus fachlicher als auch nicht-fachlicher Sicht nie Langeweile aufkam. Insbesondere möchte ich Klaus Neymeyr und Mathias Sawall nennen, welche mich stets mit Ideen und hilfreicher Kritik unterstützt haben. Für das ausführliche Korrekturlesen dieser Dissertation ist außerdem Denise Meinhardt zu danken.

Meiner Verlobten Christin Mampe danke ich für die fortwährende Unterstützung und Motivation in den vergangenen sowie allen noch kommenden Jahren.

In gleichem Maße gilt der Dank meinen Eltern Anett und Dirk, die mir stets hilfreich zur Seite stehen und mir ein unbeschwertes Studium ermöglicht haben.

Als letztes möchte ich mich bei Gordon Frank für eine langjährige Freundschaft und Begleitung durch das Mathematikstudium in Rostock bedanken.

Zusammenfassung

Das Problem der nichtnegativen Matrixfaktorisierung besitzt im Regelfall keine eindeutige Lösung. Zusatzbedingungen an die Faktoren führen auf eine reduzierte Lösungsmenge. In dieser Arbeit werden die Spalten eines Faktors durch die diskretisierte Lösung eines Anfangswertproblems eines parameterabhängigen gewöhnlichen Differentialgleichungssystems beschrieben. Dies erlaubt eine sehr effiziente Reduktion der Lösungsmenge des Faktorisierungsproblems. Es wird eine Darstellung dieser Lösungen im Raum der Parametrierungen des Anfangswertproblems eingeführt und analysiert. Zur numerischen Approximation der Lösungsmengen wird ein adaptiver, parallelisierbarer Algorithmus präsentiert. Ein Anwendungsszenario dieser Theorie findet sich in der chemometrischen Analyse von Spektrenfolgen zu chemischen Reaktionssystemen.

Abstract

The nonnnegative matrix factorization problem usually has no unique solution. Additional constraints to the factors result in a reduced solution set. In this paper, the columns of one of the factors are described by the discretized solution of an initial value problem of a parameter-dependent system of ordinary differential equations. This makes a very efficient reduction of the solution set of the factorization problem prossible. A representation of these solutions in the space of parameters of the initial value problem is introduced and analyzed. For the numerical approximation of the solution sets an adaptive, parallelizable algorithm is presented. An application scenario of this theory is the chemometric analysis of spectra sequences for chemical reaction systems.

Inhaltsverzeichnis

1. Einleitung

In der numerischen linearen Algebra und numerischen Mathematik sind Verfahren zur Faktorisierung von Matrizen von größter Bedeutung. Nahezu alle Verfahren zur Simulation oder Modellierung naturwissenschaftlicher Prozesse nutzen zumindest in einem ihrer Teilschritte ein solches Verfahren zur Zerlegung einer Matrix in zwei oder mehr Faktoren. Besonders prominente Beispiele sind die LR-Faktorisierung im Rahmen des Gaußschen Eliminationsverfahrens, die QR-Faktorisierung zur Lösung von Ausgleichsproblemen, die Eigenwertzerlegung $A = U\Lambda U^T$ und die Singulärwertzerlegung $A = U\Sigma V^T$.

In Algorithmen für die Signalverarbeitung, Datenanalyse, Data Mining und maschinelles Lernen gewinnen Faktorisierungstechniken für nichtnegative Matrizen und Tensoren zunehmend an Bedeutung [10, 88, 92, 129]. Solche Verfahren werden oft auf große experimentell gewonnene Datensätze nichtnegativer Messdaten angewandt, um Strukturinformationen von häufig niedrigdimensionaler Natur aus den hochdimensionalen Daten zu extrahieren. Für das nichtnegative Matrixfaktorisierungsproblem, im Englischen *nonnegative matrix factorization problem* (NMF-Problem), treten bei dessen Lösung im Vergleich zu den oben genannten Matrixfaktorisierungen besondere Schwierigkeiten auf. Während die LR-Faktorisierung, die QR-Faktorisierung und die Eigenwert- sowie Singulärwertzerlegung unter geeigneten Normierungs- und Orientierungsannahmen im Regelfall (abgesehen etwa von entarteten Eigenräumen) eindeutige Lösungen besitzen, gilt dies für das NMF-Problem, sofern überhaupt eine Lösung existiert, nicht. Vielmehr ist von Kontinua möglicher Faktorisierungen auszugehen. Das NMF-Problem und seine Lösungsmenge stehen im Mittelpunkt dieser Arbeit. Zusätzliche Strukturen eines Anwendungsfeldes, nämlich die nichtnegative Matrixfaktorisierung von spektroskopischen Daten, liefern die Basis für die Entwicklung von Techniken zur Identifizierung spezieller NMFs in der Menge aller NMFs einer gegebenen Matrix.

Ausgangspunkt der Untersuchungen in dieser Arbeit ist eine nichtnegative $m \times n$ Matrix D vom Rang s, deren Faktoren $C \in \mathbb{R}^{m\times s}$ und $S \in \mathbb{R}^{n\times s}$ gesucht sind, sodass $D = CS^T$ gilt. Diese Aufgabenstellung wird nichtnegative Matrixfaktorisierung (im Sinne einer Vollrangfaktorisierung) genannt und ist ein sogenanntes schlecht gestelltes inverses Problem [36, 55].

Durch die an den Rang s von D gekoppelten Dimensionsforderungen an C und S werden triviale Lösungen wie $D = DI$ mit der $n \times n$ Einheitsmatrix I ausgeschlossen, sofern $s < \min(m, n)$. Für das Problem der nichtnegativen Matrixfaktorisierung kann die Existenz einer Lösung nicht vorausgesetzt werden, wie das Beispiel der 4×4 Matrix

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{1.1}$$

zeigt [21, 123]. Diese Matrix ist vom Rang 3 und besitzt nur NMFs mit Matrixfaktoren vom Rang 4. Es wird nun angenommen, dass eine nichtnegative Matrixfaktorisierung von

D existiert. Durch die verallgemeinerten Permutationsmatrizen $P\Delta$, das sind Produkte von Permutationsmatrizen P mit nichtnegativen regulären Diagonalmatrizen Δ, lassen sich stets in trivialer Weise weitere nichtnegative Matrixfaktorisierungen generieren, denn es gilt

$$D = CS^T = C(P\Delta)^{-1}P\Delta S^T = \underbrace{(C\Delta^{-1}P^T)}_{\geq 0}\underbrace{(P\Delta S^T)}_{\geq 0} .$$

Solche auf Spaltenumordnung und -skalierung basierenden Lösungen sind nicht von Interesse. Vielmehr gibt es oft Mengen möglicher Faktorisierungen, die sich darüber hinaus wesentlich voneinander unterscheiden.

Je nach zugrunde liegender Anwendung ist zu unterscheiden, ob das Ziel in der Bestimmung einer beliebigen nichtnegativen Matrixfaktorisierung besteht oder aber eine Faktorisierung $D = \tilde{C}\tilde{S}^T$ zu identifizieren ist, welche im Aufgabenkontext zusätzliche Bedingungen erfüllt. In dieser Arbeit liegt der Fokus auf dem letzteren Fall. Da von einem Kontinuum an möglichen nichtnegativen Matrixfaktorisierungen von D ausgegangen werden muss, ist die Bestimmung von $\tilde{C}$ und $\tilde{S}$ nur dann möglich, wenn zwischen zwei unterschiedlichen Faktorisierungen innerhalb dieses Kontinuums differenziert werden kann. Hierzu können neben der Nichtnegativität weitere Bedingungen an eine Faktorisierung von D gestellt werden. Eine so erzielte Reduktion der Lösungsmenge kann als Regularisierung des Problems der nichtnegativen Matrixfaktorisierung verstanden werden. Im Idealfall besteht die reduzierte Lösungsmenge nur noch aus der Faktorisierung $\tilde{C}\tilde{S}^T$.

In dieser Arbeit stehen modellbasierte Nebenbedingungen im Fokus. Sei $C^{\mathrm{M}}(k) \in \mathbb{R}^{m\times s}$ hierzu die Auswertung eines diskretisierten Modells des Faktors C mit dem Parameter $k \in \mathbb{R}^q$. Es sind dann nur solche nichtnegativen Matrixfaktorisierungen $D = CS^T$ von Interesse, die konsistent mit dem vorgegebenen Modell sind, das heißt für die eine Parametrierung k existiert mit

$$C = C^{\mathrm{M}}(k) \quad \text{beziehungsweise} \quad D = C^{\mathrm{M}}(k)S^T .$$

Ob eine solche Regularisierung die eindeutige Bestimmung von $\tilde{C}\tilde{S}$ ermöglicht oder weitere mit dem Modell konsistente Faktorisierungen von D existieren, wird untersucht.

In einer Vielzahl von Anwendungsproblemen ist die nichtnegative Matrixfaktorisierung ein zentraler Schritt bei der Bestimmung von Lösungen. Für diese Arbeit von besonderer Bedeutung ist die Analyse von experimentell gemessenen Spektrenfolgen eines Reaktionssystems aus s Komponenten. Die so ermittelten Daten bezüglich eines (Zeit×Frequenz)-Gitters lassen sich in Form einer nichtnegativen Matrix D abspeichern. Für eine nichtnegative Matrixfaktorisierung $D = CS^T$ können die Spalten von C den Konzentrationsprofilen und die Spalten von S den sogenannten Reinspektren der s Komponenten zugeordnet werden. Es existiert (abgesehen von Umordnungen der Spalten der Faktoren) genau eine Faktorisierung $D = \tilde{C}\tilde{S}^T$, die den chemischen Sachverhalt korrekt wiedergibt. Die Bestimmung dieser Faktorisierung lässt sich oft nur durch den Einsatz geeigneter Nebenbedingungen umsetzen. In dieser Arbeit sind dies sogenannte kinetische Modelle, die näherungsweise die Konzentrationsprofile in den Spalten von C abbilden. Sie basieren auf parameterabhängigen Anfangswertproblemen der Form

$$\frac{d}{dt}c(t) = M(k)c(t), \quad c(0) = c_0 \tag{1.2}$$

mit $c : \mathbb{R} \to \mathbb{R}^s$, $k \in \mathbb{R}^q$ und $M(k) \in \mathbb{R}^{s\times s}$. Die Matrix $C^{\mathrm{M}}(k)$ zur Modellierung des Faktors C resultiert aus der Auswertung der Lösung von (1.2) zum gegebenen Zeitgitter und bekannten Anfangswerten c_0.

Das durch (1.2) regularisierte NMF-Problem besitzt nicht selten eine nichttriviale Lösungsmenge. In dieser Arbeit wird gezeigt, dass diese sich im Raum der Parameter k des Anfangswertproblems (1.2) darstellen lässt. Weiter werden grundlegende Eigenschaften solcher Lösungsmengen hergeleitet und Gleichungen zu deren Beschreibung für ausgewählte Anfangswertprobleme formuliert. Zur numerischen Approximation dieser Parameterlösungsmengen wird ein adaptiver, parallelisierbarer Algorithmus präsentiert. Alle Betrachtungen erfolgen sowohl für idealisierte als auch für störungsbehaftete Ausgangsdaten D. Die Anwendbarkeit der theoretischen Ergebnisse wird anhand von verschiedenen spektroskopisch vermessenen Reaktionssystemen demonstriert.

Grundlegende Einführungen in die Theorie der nichtnegativen Matrixfaktorisierung sind in [8, 11, 123] zu finden. Eine Beschreibung entsprechender numerischer Berechnungsmethoden erfolgt in [66, 77] und weitere Anwendungen lassen sich unter anderem in der Astronomie [117], Biologie [38], Chemie [70], Geographie [131] und Informatik [10] finden. Eine sehr umfangreiche Übersicht von Beispielen wird in [42] präsentiert. Die Analyse spektroskopischer Messserien unter Verwendung von Matrixfaktorisierungen wird in [81, 82] diskutiert. Weiterführend sind detaillierte Untersuchungen der entsprechenden Lösungsmengen in [74, 101, 102, 123] zu finden. Auf die Regularisierung des Problems der nichtnegativen Matrixfaktorisierung durch parameterabhängige kinetische Modelle wird in [26, 81] eingegangen. Im Rahmen dieser Dissertation werden insbesondere Ansätze auf Grundlage der Singulärwertzerlegung [49] diskutiert. Die Analyse von nichtnegativen Matrixfaktorisierungen, die Konsistenz zu einem kinetischen Modell aufweisen, bilden den Kern dieser Dissertation. Die Frage nach der Eindeutigkeit entsprechender Modellparametrierungen wird allgemein durch den Begriff der Identifizierbarkeit beschrieben [44, 128]. Im Kontext spektroskopischer Messreihen und kinetischer Modelle sind [125, 126] zentrale Arbeiten.

Aufbau der Arbeit

In Kapitel 2 wird das Problem der nichtnegativen Matrixfaktorisierung sowie geeignete Verallgemeinerungen für störungsbehaftete Daten und die Verwendung von weiteren Nebenbedingungen für die Faktoren betrachtet. Es werden geeignete Darstellungen der intrinsischen Lösungsuneindeutigkeit beschrieben und numerische Verfahren zur Bestimmung entsprechender Lösungen diskutiert. Die Einführung kinetischer Modelle erfolgt in Kapitel 3. Es werden Erweiterungen der numerischen Methoden zur Lösung des regularisierten nichtnegativen Faktorisierungsproblems durchgeführt und grundlegende Eigenschaften kinetischer Modelle hergeleitet. Den Kern der Arbeit bildet Kapitel 4, in dem die Parameterlösungsmengen des durch (1.2) regularisierten NMF-Problems eingeführt und ausführlich analysiert werden. Darauf aufbauend werden in Kapitel 5 Verallgemeinerungen der Parameterlösungsmengen für störungsbehaftete Ausgangsdaten D präsentiert. Für die praktische Anwendung ist eine effiziente und robuste Berechnung dieser Lösungsmengen von Parametrierungen wichtig. In Kapitel 6 wird eine entsprechende numerische Methode zu deren Approximation präsentiert. Durch Beispiele aus den Bereichen der homogenen Katalyse und der Photochemie wird in Kapitel 7 die Anwendbarkeit der hergeleiteten Methoden auf spektroskopische Datensätze demonstriert. Abschließend wird

in Kapitel 8 ein Ausblick auf fortführende Forschungsthemen gegeben.

Notationen

$\mathbb{R}_+$	Menge der strikt positiven reellen Zahlen
$\mathbb{R}_{\geq 0}$	Menge der nichtnegativen reellen Zahlen
$\mathrm{rg}(D)$	Rang der Matrix D
$\mathrm{tr}(D)$	Spur der Matrix D
$D_{i,:} = D(i,:)$	i-te Zeile von D
$D_{:,j} = D(:,j)$	j-te Spalte von D
$D_{I,:}$	Untermatrix von D, die durch Streichen aller Zeilen, deren Index nicht Element von I ist, entsteht
$s \in \mathbb{N}$	Spezies-/Komponentenanzahl
$D = CS^T$	Matrixfaktorisierung mit $D \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{m\times s}, S \in \mathbb{R}^{n\times s}$
$\phi, k \in \mathbb{R}_+^q$	Geschwindigkeitsparameter eines kinetischen Modells
$M, M(k), M(\phi)$	Koeffizientenmatrizen eines Differentialgleichungssystems eines kinetischen Modells
$\mathcal{K}$	Menge D-konsistenter Parameter
$\mathcal{K}^+$	Menge zulässiger Parameter
$\mathcal{K}_\varepsilon$	Menge D-approximativer Parameter
$\mathcal{K}^+_{\varepsilon,\theta}$	Menge zulässiger D-approximativer Parameter
$\mathcal{T}$	Operator zur Auswertung einer Funktion $c(t) : \mathbb{R} \to \mathbb{R}^s$ auf einem Gitter

2. Nichtnegative Matrixfaktorisierungen

In diesem Kapitel wird das Problem der nichtnegativen Matrixfaktorisierung betrachtet. Seit ihrer Einführung durch Paatero und Tapper [89] im Jahr 1994 und verstärkt durch die Beschreibung entsprechender Berechnungsmethoden durch Lee und Seung [76] kann ein stetig wachsendes Interesse an diesem Forschungsfeld festgestellt werden. Abbildung 2.1 veranschaulicht hierzu die Entwicklung der Anzahl von Veröffentlichungen mit ausgewählten Schlagwörtern.

In den folgenden Abschnitten werden aufbauend auf der nichtnegativen Matrixfaktorisierung drei modifizierte Aufgabenstellungen definiert, eine niedrigdimensionale Darstellung der zugehörigen Lösungsmengen beschrieben und numerische Lösungsansätze erläutert. Abschließend werden Anwendungen der nichtnegativen Matrixfaktorisierung betrachtet, wobei der Analyse spektroskopischer Messfolgen für diese Arbeit eine besondere Bedeutung zukommt.

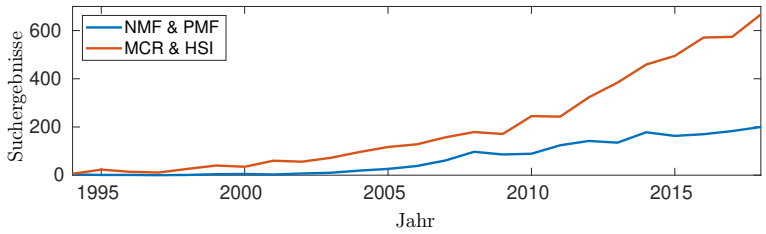


Abbildung 2.1.: Ergebnisanzahl für Google-Scholar-Suchen nach Veröffentlichungen mit Titeln, die die Stichwörter *nonnegative matrix factorization* (NMF) und *positive matrix factorization* (PMF) beziehungsweise *multivariate curve resolution* (MCR) und *hyperspectral imaging* (HSI) enthalten. Hierbei können NMF/PMF eher mit Veröffentlichungen zur Theorie und MCR/HSI vorrangig mit Anwendungen assoziiert werden.

2.1. Faktorisierungsaufgaben

Aufbauend auf der Definition der nichtnegativen Matrixfaktorisierung werden nun die, für diese Arbeit relevanten, Faktorisierungsaufgaben definiert. Anschließend werden die Existenz und die Eindeutigkeit entsprechender Lösungen kurz diskutiert. Die folgenden Definitionen entsprechen weitestgehend der Notation in [101]. Für Übersichtsarbeiten zur nichtnegativen Matrixfaktorisierung wird auf [8, 11, 123] verwiesen.

Problem 1 (Nichtnegative Matrixfaktorisierung)**.** *Gegeben sei eine nichtnegative Matrix* $D \in \mathbb{R}^{m \times n}$*. Es sind nichtnegative Matrizen* $C \in \mathbb{R}^{m \times s}$ *und* $S \in \mathbb{R}^{n \times s}$ *zu minimalem* $s \in \mathbb{N}$ *zu bestimmen, sodass* $D = CS^T$ *gilt.*

Die Fragestellung nach dem kleinstmöglichen s, sodass C und S existieren, führt auf den nichtnegativen Rang [21, 51].

Definition (Nichtnegativer Rang)**.** Gegeben sei eine nichtnegative Matrix $D \in \mathbb{R}^{m\times n}$. Die minimale Zahl $r \in \mathbb{N}$, für die nichtnegative Faktoren $C \in \mathbb{R}^{m\times r}$ und $S \in \mathbb{R}^{n\times r}$ mit $D = CS^T$ existieren, heißt nichtnegativer Rang $\mathrm{rg}_+(D)$.

Häufig ist zudem von Interesse, ob auch zu einem vorgegeben s eine Faktorisierung bestimmt werden kann. Es ist leicht nachzuvollziehen, dass Problem 1 für $s < \mathrm{rg}(D)$ keine Lösung besitzt. Aus dem Grenzfall $s = \mathrm{rg}(D)$ geht die folgende Fragestellung hervor:

Problem 2 (Nichtnegative Vollrangfaktorisierung)**.** *Gegeben sei die nichtnegative Matrix $D \in \mathbb{R}^{m\times n}$ mit $s := \mathrm{rg}(D)$. Es sind nichtnegative Matrizen $C \in \mathbb{R}^{m\times s}$ und $S \in \mathbb{R}^{n\times s}$ zu bestimmen, sodass $D = CS^T$ gilt.*

Die nichtnegative Vollrangfaktorisierung besitzt also eine Lösung, wenn $s = \mathrm{rg}(D) = \mathrm{rg}_+(D)$ gilt, vergleiche [21]. Es folgt nun eine Verallgemeinerung der nichtnegativen Vollrangfaktorisierung unter der Annahme gestörter Ausgangsdaten D. Darunter ist zu verstehen, dass D betragskleine negative Einträge enthalten kann und statt $s = \mathrm{rg}(D)$ nun $s < \mathrm{rg}(D)$ gilt.

Problem 3 (Approximative (nichtnegative) Matrixfaktorisierung)**.** *Gegeben seien eine Matrix $D \in \mathbb{R}^{m\times n}$ mit $D_{i,j} \gg -\max_{k,l} |D_{k,l}|$ für $i = 1,\dots,m,\ j = 1,\dots,n$, ein $s \in \mathbb{N}$ mit $s < \mathrm{rg}(D)$ und ein $\varepsilon \geq 0$. Es sind Matrizen $C \in \mathbb{R}^{m\times s}$ und $S \in \mathbb{R}^{n\times s}$ mit*

$$\min_{i=1,\dots,m} C_{i,l} \Big/ \max_{i=1,\dots,m} C_{i,l} \geq -\varepsilon \ \text{ und } \ \min_{i=1,\dots,n} S_{i,l} \Big/ \max_{i=1,\dots,n} S_{i,l} \geq -\varepsilon \ \text{ für } l = 1,\dots,s \quad (2.1)$$

zu bestimmen, die $\|D - CS^T\|_F$ minimieren.

Besteht das Ziel eines Anwendungsproblems nicht in der Bestimmung einer beliebigen nichtnegativen beziehungsweise approximativen Faktorisierung, sondern einer solchen mit speziellen Eigenschaften, können weitere Nebenbedingungen an die Faktoren C und S gestellt werden. Dies entspricht einer Regularisierung des entsprechenden Faktorisierungsproblems.

Problem 4 (Regularisierte Matrixfaktorisierung)**.** *Seien die Voraussetzungen wie in Problem 3 und eine Funktion $g : \mathbb{R}^{m\times s} \times \mathbb{R}^{n\times s} \to \mathbb{R}_{\geq 0}$ gegeben. Es sind Matrizen $C \in \mathbb{R}^{m\times s}$ und $S \in \mathbb{R}^{n\times s}$ zu bestimmen, die* (2.1) *erfüllen und $\|D - CS^T\|_F + g(C,S)$ minimieren.*

Die eingeführte regularisierte Matrixfaktorisierung stellt eine Grundform dieses Problemtyps dar. Eine Regularisierung kann beispielsweise auch nur von C, nur von S oder aber von weiteren Parametern abhängen. Auf die Definition dieser Spezialfälle wird verzichtet. Siehe dazu etwa [87, 100].

Existenz und Eindeutigkeit von Lösungen

Bei den vier vorgestellten Faktorisierungsaufgaben handelt es sich um inverse Probleme [36]. Weiter heißt ein inverses Problem nach [55] *korrekt gestellt*, wenn es die drei folgenden Eigenschaften erfüllt:

1. es existiert eine Lösung,
2. die Lösung ist eindeutig bestimmt und
3. sie hängt stetig von den Ausgangsdaten ab.

Ist eine der drei Eigenschaften nicht erfüllt, heißt das Problem *schlecht gestellt.*

Das Problem 1 besitzt für die Einheitsmatrizen $I_m \in \mathbb{R}^{m\times m}$ und $I_n \in \mathbb{R}^{n\times n}$ immer Faktorisierungen der Form $D = DI_n = I_m D$. Hierbei ist klar, dass $s \in \{m, n\}$ nicht zwangsläufig minimal ist. Dennoch folgt daraus, dass immer eine (gegebenenfalls triviale) Lösung existiert. Die Existenz einer Lösung kann für die nichtnegative Vollrangfaktorisierung nicht mehr vorausgesetzt werden. Die 4×4 Matrix aus (1.1) im einleitenden Kapitel verdeutlicht dies, denn ihr Rang ist 3, aber es existiert keine nichtnegative Vollrangfaktorisierung mit $s = 3$, siehe [21,123]. Weiter lässt sich leicht für die Probleme 1 und 2 zeigen, dass eine Lösung keineswegs eindeutig sein muss. Sei hierzu eine nichtnegative Matrix D mit $\mathrm{rg}(D) = \mathrm{rg}_+(D)$ gegeben. Damit existiert für $s = \mathrm{rg}(D)$ eine nichtnegative Faktorisierung CS^T, die beide Probleme löst. Für verallgemeinerte Permutationsmatrizen $P\Delta \in \mathbb{R}^{s\times s}$ und $P\Delta \neq I$ ist durch $\tilde{C}\tilde{S}^T = (C(P\Delta)^{-1})(P\Delta\tilde{S}^T)$ eine weitere Lösung der Probleme 1 und 2 gegeben. Bei beiden Faktorisierungsaufgaben handelt es sich also um schlecht gestellte inverse Probleme. Für weiterführende Betrachtungen zur Eindeutigkeit der nichtnegativen Matrix- und Vollrangfaktorisierung sei auf [73, 74, 123] verwiesen.

Diese Betrachtungen lassen sich wegen der betragskleinen negativen Einträge nicht direkt auf die Probleme 3 und 4 übertragen. Im Allgemeinen kann aber auch hier analog zur nichtnegativen Vollrangfaktorisierung weder von der Existenz noch der Eindeutigkeit einer Lösung ausgegangen werden.

2.2. Matrixfaktorisierung mittels Singulärwertzerlegung

Werden die Einträge der Faktoren einer nichtnegativen Vollrangfaktorisierung als Freiheitsgrade betrachtet, sind diese im Allgemeinen stark redundant. Mittels der Singulärwertzerlegung kann die Anzahl dieser Freiheitsgrade signifikant reduziert werden.

Satz 1 (Singulärwertzerlegung, [49])**.** *Sei ein $D \in \mathbb{R}^{m\times n}$ mit $r = \mathrm{rg}(D)$ gegeben. Es existieren orthogonale Matrizen $U \in \mathbb{R}^{m\times m}$ und $V \in \mathbb{R}^{n\times n}$ sowie eine Matrix $\Sigma \in \mathbb{R}^{m\times n}$, deren einzige Nichtnulleinträge durch $\Sigma_{i,i} = \sigma_i$, $i = 1, \ldots, r$, mit $\sigma_1 \geq \ldots \geq \sigma_r > 0$ definiert sind, sodass*

$$U^T D V = \Sigma\,.$$

Hierbei werden $U\Sigma V^T$ die Singulärwertzerlegung von D und σ_i für $i = 1, \ldots, r$ die Singulärwerte der Matrix D genannt. Die Spalten der Matrizen $U = [u_1, \ldots, u_m]$ und $V = [v_1, \ldots, v_n]$ heißen links- beziehungsweise rechtsseitige Singulärvektoren.

Satz 2 (Eckard-Young Theorem, [34])**.** *Seien $D \in \mathbb{R}^{m\times n}$, $p = \min(m, n)$, $s \in \mathbb{N}$ mit $0 \leq s \leq p$ und eine Singulärwertzerlegung $D = U\Sigma V^T = \sum_{i=1}^{p} \sigma_i u_i v_i^T$ von D gegeben. Für die Matrix $D_s := \sum_{i=1}^{s} \sigma_i u_i v_i^T$ gilt*

$$\|D - D_s\|_2 = \min_{\substack{A\in\mathbb{R}^{m\times n}\\ \mathrm{rg}(A)\leq s}} \|D - A\|_2 = \sigma_{s+1} \quad \textit{mit} \quad \sigma_{s+1} := 0 \textit{ für } s = p,$$

$$\|D - D_s\|_F = \min_{\substack{A\in\mathbb{R}^{m\times n}\\ \mathrm{rg}(A)\leq s}} \|D - A\|_F = \sqrt{\sigma_{s+1}^2 + \ldots + \sigma_r^2}.$$

Aus den Sätzen 1 und 2 folgt, dass eine Matrix D vom Rang s gleich der Matrix D_s ist, sie also bereits durch $U(1:s,:)$, $V(1:s,:)$ und $\Sigma(1:s,1:s)$ exakt rekonstruiert wird.

Der vereinfachten Notation halber seien die Faktoren einer solchen abgeschnittenen Singulärwertzerlegung im Folgenden stets mit U, Σ und V bezeichnet. Für eine nichtnegative Matrix D mit $s = \mathrm{rg}(D) = \mathrm{rg}_+(D)$ lässt sich durch eine solche abgeschnittene Singulärwertzerlegung mit $D = (U\Sigma)V^T$ eine Faktorisierung bestimmen, deren Faktoren $U\Sigma$ und V bereits die gewünschte Anzahl s an Spalten haben, aber eventuell negative Einträge besitzen. Um auch die Nichtnegativitätsbedingungen von Problem 2 zu erfüllen, werden geeignete Linearkombinationen der Spalten von $(U\Sigma)$ und V betrachtet. Dies geschieht nach [47, 81] durch Einführung einer reguläre Matrix $T \in \mathbb{R}^{s\times s}$, sodass

$$D = U\Sigma V^T = \underbrace{U\Sigma T^{-1}}_{C}\underbrace{TV^T}_{S^T}. \tag{2.2}$$

Dass sich jede Lösung der nichtnegativen Vollrangfaktorisierung $D = CS^T$ stets wie in (2.2) mit einem geeigneten T darstellen lässt, ist in [87] bewiesen. Dadurch ergibt sich für Matrizen $D \in \mathbb{R}^{m\times n}$ und $s \ll m, n$ eine signifikante Reduktion der Anzahl an Freiheitsgraden. Lautet die Anzahl an Komponenten in den Faktoren C und S zusammen noch $s(m+n)$, sind es für die Matrix T nur s^2. Eine Konsequenz aus diesen Überlegungen ist, dass sich die Faktorisierungsaufgaben unter Nutzung der Singulärwertzerlegung auf die Bestimmung geeigneter Matrizen T reduzieren lassen.

2.3. Menge zulässiger Lösungen

In diesem Abschnitt wird für Matrizen D mit $s = \mathrm{rg}(D) = \mathrm{rg}_+(D)$ und der entsprechenden nichtnegativen Vollrangfaktorisierung $D = CS^T$ eine Projektion der Spalten des Faktors S in einen $(s-1)$-dimensionalen Raum erläutert. Dieser ermöglicht zudem für $s \leq 4$ die grafische Darstellung der Lösungsmenge der genannten Faktorisierungsaufgabe. Alle Überlegungen sind analog unter Betrachtung der Transposition $D^T = SC^T$ auch für C anwendbar.

Festlegung einer Skalierung der Spalten von S

In Abschnitt 2.1 wurde bereits festgestellt, dass eine Skalierung der Spalten der Faktoren C und S durch positive reelle Zahlen deren Nichtnegativität nicht beeinflusst. Es wird im Folgenden davon ausgegangen, dass ein solches Umskalieren in nichtnegativen Vollrangfaktorisierungen resultiert, die lediglich redundante Informationen enthalten. Es wird eine Möglichkeit erläutert, dies für eine weitere Reduktion der Freiheitsgrade auf Basis der Perron-Frobenius-Theorie [40, 85, 93] zu nutzen. Aufbauend folgt eine Aussage zur Irreduzibilität von DD^T.

Lemma 1. *Sei $D \in \mathbb{R}^{m\times n}$ eine nichtnegative Matrix mit $m \geq 2$, welche keine Nullzeilen enthält. Es ist DD^T genau dann reduzibel, wenn es Permutationsmatrizen $P \in \mathbb{R}^{m\times m}$ und $Q \in \mathbb{R}^{n\times n}$ gibt, sodass*

$$PDQ = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

mit $D_1 \in \mathbb{R}^{m'\times n'}$ und $D_2 \in \mathbb{R}^{(m-m')\times(n-n')}$ sowie $0 < m' < m$ und $0 < n' < n$ ist.

Beweis: Siehe [101]. □

Die Voraussetzung, dass keine Nullzeilen in D vorhanden sind, kann für Anwendungen typischerweise durch Streichen entsprechender Zeilen erreicht werden, weil diese für eine

nichtnegative Vollrangfaktorisierung in C ebenfalls Nullzeilen generieren [101]. Dies folgt analog für Nullspalten. Durch sukzessive Anwendung von Lemma 1 kann angenommen werden, dass D durch geeignete Zeilen- und Spaltenpermutationen in eine Blockdiagonalform überführt werden kann, sodass für jeden dieser Blöcke B die Matrizen BB^T und B^TB irreduzibel sind. Auch die Faktorisierungsaufgaben gehen in eine entsprechende Blockstruktur über. Statt der Matrix D können auch die entsprechenden Unterblöcke betrachtet werden. Ohne Beschränkung der Allgemeinheit kann also stets von Matrizen D ausgegangen werden, für die DD^T und D^TD irreduzibel sind. Die Perron-Frobenius-Theorie besagt nun unter anderem, dass für die nichtnegative, irreduzible Matrix DD^T der Eigenvektor zum größten Eigenwert komponentenweise strikt positiv ist. Dies gilt analog für D^TD und resultiert für die Singulärwertzerlegung $D = U\Sigma V^T$ in entweder strikt positiven oder strikt negativen ersten links- und rechtsseitigen Singulärvektoren, also $U(:,1) > 0, V(:,1) > 0$ oder $U(:,1) < 0, V(:,1) < 0$. Durch geeignete Multiplikation mit -1 kann die strikte Positivität angenommen werden. Darauf aufbauend stellt das folgende Resultat den finalen Schritt für die angestrebte Skalierung der Spalten von S beziehungsweise der Zeilen von T dar.

Lemma 2. *Seien eine nichtnegative Matrix $D \in \mathbb{R}^{m\times n}$ mit $s = \mathrm{rg}(D) \geq 2$ mit einer Singulärwertzerlegung $D = U\Sigma V^T$ gegeben. Ist D^TD irreduzibel, so gilt für jede nichtnegative Linearkombination $Vx \geq 0$ mit $x \in \mathbb{R}^s$ und $\|x\| > 0$, dass $x_1 \neq 0$ ist.*

Beweis: Siehe [86]. □

Aus Lemma 2 folgt für die Matrix T in (2.2), dass $T_{i,1} \neq 0$ für $i = 1, \ldots, s$ gilt. Folglich lässt sich jede Zeile von T so skalieren, dass der jeweils erste Eintrag der resultierenden Zeilen 1 ist. Ist die Skalierung der Spalten von C und S frei wählbar, so kann sich also im Folgenden auf Matrizen T der Form

$$T = \begin{pmatrix} 1 & t_{1,2} & \cdots & t_{1,s} \\ \vdots & \vdots & & \vdots \\ 1 & t_{s,2} & \cdots & t_{s,s} \end{pmatrix} \tag{2.3}$$

beschränkt werden. Die Darstellung der Spalten von S erfolgt hier also im $(s-1)$-dimensionalen Raum der Expansionskoeffizienten der rechtsseitigen Singulärvektoren zu den Singulärwerten σ_2 bis σ_s. Die s Zeilen von T zur vollständigen Darstellung einer nichtnegativen Vollrangfaktorisierung entsprechen also s Punkten in diesem Raum.

Permutation der Spalten von S

Abschließend wird in diesem Abschnitt durch Betrachtung der möglichen Permutationen der Spalten von S eine Darstellung der Lösungsmenge der nichtnegativen Vollrangfaktorisierung eingeführt. Analog zur Skalierung ist die Permutation der Spalten von S gleichbedeutend mit der (selben) Permutation der Zeilen von T. Auch eine solche Umsortierung der Zeilen von T und den daraus folgenden Lösungen der nichtnegativen Vollrangfaktorisierung werden als redundante Informationen angesehen. Folglich reicht es, nur die erste Zeile von T zu betrachten und für den Rest eine geeignete Ergänzung zu fordern. Mit $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^{s-1}$ ist die *Menge zulässiger Lösungen* zum Faktor S durch

$$\mathcal{M}_S := \left\{ t \in \mathbb{R}^{s-1} : \exists T = \begin{pmatrix} 1 & t^T \\ \mathbf{1} & * \end{pmatrix} \text{ mit } \mathrm{rg}(T) = s, U\Sigma T^{-1} \geq 0, TV^T \geq 0 \right\}$$

definiert. Damit ist $\mathcal{M}_S$ eine $(s-1)$-dimensionale Darstellung aller möglichen Spalten des Faktors S. Die Menge $\mathcal{M}_C$ zur Darstellung aller möglichen Spalten des Faktors C folgt analog aus der Betrachtung der Transposition $D^T = SC^T$. Eine nichtnegative Vollrangfaktorisierung lässt sich durch s Elemente der Mengen $\mathcal{M}_S$ oder $\mathcal{M}_C$ darstellen. Die Umkehrung, inwiefern sich aus s Elementen aus $\mathcal{M}_S$ oder $\mathcal{M}_C$ eine entsprechende Faktorisierung ergibt, wird in [65, 101, 103, 106] diskutiert.

Abschließend sei auszugsweise auf weiterführende Literatur hingewiesen. Die Lösungsmengen nichtnegativer Vollrangfaktorisierungen für $s = 2$ wurden 1971 durch Lawton und Sylvestre [75] analysiert und bestimmt. Eine entsprechende Erweiterung zu einem auf Dreieckskonstruktionen basierenden Ansatz für die Bestimmung von $\mathcal{M}_S$ mit $s = 3$ erfolgte 1985 durch Borgen und Kowalski [13]. Darüber hinaus sei für detaillierte Analysen der Menge zulässiger Lösungen mit $s \leq 4$ auf [46, 95, 106] verwiesen. Numerische Algorithmen zur Approximation von $\mathcal{M}_S$ beziehungsweise $\mathcal{M}_C$ sind unter anderem *grid search* [127], *triangle enclosure* [47], *polygon/polyhydron inflation* [102, 105], *particle swarm optimization* [115], *ray casting* [108] und die *simultane Konstruktion dualer Borgen-Plots* [106]. Weiter wird darauf hingewiesen, dass für $s = 3$ auch Mengen zulässiger Lösungen mit einem Segment [107] oder (in Spezialfällen) mit $3n, n \in \mathbb{N}$, Segmenten [110] auftreten können.

2.4. Numerische Berechnung

Es wird nun eine numerische Methode vorgestellt, die sich zur Berechnung einer beliebigen Lösung der nichtnegativen Vollrangfaktorisierung beziehungsweise der approximativen Matrixfaktorisierung eignet. Sie basiert erneut auf der Singulärwertzerlegung der Matrix D. Auszugsweise seien weitere in der Literatur zu findende Ansätze genannt [66, 76, 77, 79, 89].

2.4.1. Idealisierte Ausgangsdaten

Als Ausgangspunkt für die weiteren Betrachtungen dient erneut der Fall einer nichtnegativen Matrix $D \in \mathbb{R}^{m \times n}$ mit $s = \operatorname{rg}(D) = \operatorname{rg}_+(D)$. Es wird nun eine Minimierungsaufgabe hergeleitet, deren numerische Lösung in einer nichtnegativen Vollrangfaktorisierung resultieren kann.

Ausgehend von Gleichung (2.2) werden die, von T abhängigen, Matrizen $C := C(T) = U\Sigma T^+$ und $S := S(T) = VT^T$ eingeführt. Für den Fall des Auftretens singulärer Matrizen T wird anstelle der Inversen T^{-1} die Pseudoinverse T^+ genutzt. Es sei die Minimierungsaufgabe

$$F(T) \to \min \tag{2.4}$$

mit der Zielfunktion

$$\begin{aligned} F(T) := \gamma_1 \|I - TT^+\|_F^2 + \gamma_2 \sum_{i=1}^{m} \sum_{j=1}^{s} \left(\min \left(\frac{C_{i,j}}{\max_l(C_{l,j})}, 0 \right) \right)^2 \\ + \gamma_3 \sum_{i=1}^{n} \sum_{j=1}^{s} \left(\min \left(\frac{S_{i,j}}{\max_l(S_{l,j})}, 0 \right) \right)^2 \end{aligned} \tag{2.5}$$

zu Gewichtungsfaktoren $\gamma_1, \gamma_2, \gamma_3 \in \mathbb{R}_+$ definiert. Die drei Summanden der Gleichung (2.5) werden auch *Straffunktionen* genannt und sollen sicherstellen, dass die Anforderungen an eine nichtnegative Vollrangfaktorisierung erfüllt werden. Nach Konstruktion von (2.5) ist $F(T) \geq 0$ und es gilt $F(T) = 0$ genau dann, wenn T regulär ist und $C \geq 0$ sowie $S \geq 0$ gelten. Die Startiterierte sei im Folgenden stets als $T^{(0)}$ bezeichnet. Durch die Lösung T^* von (2.4) ist bei geeigneter Wahl von $T^{(0)}$ eine nichtnegative Vollrangfaktorisierung $D = C(T^*)(S(T^*))^T$ bestimmt. Die in Gleichung (2.3) eingeführte Form der Matrix T mit Einsen in der ersten Spalte kann analog genutzt werden, um eine Reduktion der Anzahl an zu optimierenden Freiheitsgraden in (2.4) zu erreichen. Geeignete iterative Verfahren zur Durchführung der Minimierung in Form eines Quadratmittelproblems sind beispielsweise das Gauß-Newton Verfahren und die Klasse der Quasi-Newton Verfahren [12, 67, 90]. In dieser Arbeit wird der darauf aufbauende "*trust-region-reflective*"-Algorithmus [22] in der Matlab-Implementierung *lsqnonlin* für alle Minimierungen verwendet.

2.4.2. Störungsbehaftete Ausgangsdaten

Nun wird eine Verallgemeinerung der numerischen Methode aus dem vorherigen Abschnitt 2.4.1 für Matrizen

$$\tilde{D} = D + E$$

eingeführt. Für die Matrix D gilt weiterhin $\mathrm{rg}(D) = \mathrm{rg}_+(D)$ und die Matrix $E \in \mathbb{R}^{m \times n}$ repräsentiert Störungen, wobei $|E_{i',j'}| \ll \max_{i,j} |D_{i,j}|$ für alle i', j' gilt. Es wird nun eine Minimierungsaufgabe hergeleitet, deren numerische Lösung in einer approximativen Matrixfaktorisierung resultieren kann. Hierzu werden Auswirkungen von Störungen auf die entsprechende Faktorisierungsaufgabe diskutiert und dann eine Verallgemeinerung der Zielfunktion (2.5) hergeleitet.

Auswirkungen von Störungen auf die nichtnegative Vollrangfaktorisierung

Bemerkung. Es ist davon auszugehen, dass eine geeignete Datenvorbehandlung[1] durchgeführt wurde und eine weitere Reduktion der Störungseinflüsse E nicht möglich ist. Exemplarisch für die in E zusammengefassten Störungen seien Rauschen und lokal begrenzte Störungen (Artefakte) sowie systematische globale Störungen (Grundlinienfehler) genannt. Eine Veranschaulichung folgt in Abschnitt 2.4.3 für ein Modellproblem.

Die Bestimmung einer nichtnegativen Vollrangfaktorisierung ist für $\tilde{D}$ im Allgemeinen nicht möglich. Es gilt zum einen, dass $\tilde{D}$ betragskleine negative Einträge besitzen kann, und zum anderen

$$\min(m, n) = \mathrm{rg}(\tilde{D}) = \mathrm{rg}_+(\tilde{D}) > \mathrm{rg}_+(D) = \mathrm{rg}(D)\,.$$

Hieraus resultieren zwei Schwierigkeiten:

1. Weil nur $\tilde{D}$ und nicht D bekannt ist, muss s separat gewählt oder aus $\tilde{D}$ hergeleitet werden.
2. Aus $\tilde{D}$ ist eine geeignete Approximation der Matrix D zu bestimmen, die wiederum zur Rekonstruktion der Faktorisierung $\tilde{D} \approx D = CS^T$ genutzt wird.

[1] Exemplarisch seien die Modellierung und Subtraktion einer Grundlinie [45, 97], das "asymmetric least squares smoothing" [35] und Rauschfilter [6, 99] genannt.

Bestimmung von s

Eine korrekte Wahl von s ist in Anwendungsproblemen häufig entscheidend für die Interpretierbarkeit einer Lösung der approximativen Matrixfaktorisierung. Im einfachsten Fall ist s bereits durch das zugrunde liegende Anwendungsproblem definiert. Andernfalls können zum Beispiel die Singulärwerte der Singulärwertzerlegung $\tilde{D} = U\Sigma V^T$ betrachtet werden. Es sei darauf hingewiesen, dass es sich bei diesem Schritt um ein heuristisches Vorgehen handelt. Sind in der Folge von Singulärwerten $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)}$ die Elemente σ_1 bis σ_i signifikant größer als alle weiteren Singulärwerte, ist $s = i$ häufig eine geeignete Wahl.

Bestimmung einer geeigneten Approximation von $\tilde{D}$ mit dem Rang s

Für ein festgelegtes s folgt mit Satz 2

$$\tilde{D} = USV^T = \underbrace{\sum_{i=1}^{s} \sigma_i u_i v_i^T}_{=D_s \approx D} + \underbrace{\sum_{i=s+1}^{\min(m,n)} \sigma_i u_i v_i^T}_{\approx E} .$$

Die Matrix D_s wird *Niedrigrangapproximation vom Rang s* der Matrix $\tilde{D}$ genannt und ist bezüglich der euklidischen und der Frobeniusnorm optimal. Weiter kann die Matrix D_s allerdings immer noch betragskleine negative Einträge aufweisen. Durch die Lösung der approximativen Matrixfaktorisierung von D_s ist dennoch im Allgemeinen eine gute Näherung an eine nichtnegative Vollrangfaktorisierung von D zu erwarten.

Numerische Berechnung einer approximativen Matrixfaktorisierung

Ausgehend von der Matrix D_s mit abgeschnittener Singulärwertzerlegung $D_s = U\Sigma V^T$ und $C = C(T) = U\Sigma T^+$ sowie $S = S(T) = VT^T$, wie in Abschnitt 2.4.1, kann die Zielfunktion

$$\begin{aligned} F(T) :=& \gamma_1 \|I - TT^+\|_F^2 + \gamma_2 \sum_{i=1}^{m} \sum_{j=1}^{s} \left(\min\left(\frac{C_{i,j}}{\max_l(C_{l,j})} + \varepsilon, 0 \right) \right)^2 \\ &+ \gamma_3 \sum_{i=1}^{n} \sum_{j=1}^{s} \left(\min\left(\frac{S_{i,j}}{\max_l(S_{l,j})} + \varepsilon, 0 \right) \right)^2 \end{aligned} \tag{2.6}$$

mit der Fehlertoleranz $\varepsilon \geq 0$ definiert werden. Für den Fall $\varepsilon = 0$ entspricht (2.6) der Zielfunktion (2.5), was die identische Bezeichnung rechtfertigt.

2.4.3. Anwendung der numerischen Methoden für ein Modellproblem

Anhand der folgenden Modellprobleme werden die vorgestellten numerischen Berechnungsmethoden für die nichtnegative Vollrangfaktorisierung und die approximative Matrixfaktorisierung demonstriert.

Modelldatensatz 1. Seien die Vektoren $x = (0,\ 0.05, \ldots, 5)^T$ und $y = (0,\ 0.01, \ldots, 4)^T$ sowie die Funktionen

$$f_1(\xi) = \mathrm{e}^{-\xi}, \quad f_2(\xi) = \mathrm{e}^{-\xi} - \mathrm{e}^{-2\xi}, \quad f_3(\xi) = 1 - f_1(\xi) - f_2(\xi) \quad \text{und}$$

$$h_k(\mu) = 0.9 \cdot \mathrm{e}^{\frac{-(\mu-k)^2}{0.3}} + 0.1 \cdot \mathrm{e}^{\frac{-(\mu-k)^2}{10}}, \quad k = 1, 2, 3 \ ,$$

gegeben.

Hierdurch sind mit $C_{i,k} = f_k(x_i)$, $i = 1, \ldots, 101$, und $S_{j,k} = h_k(y_j)$, $j = 1, \ldots, 401$, für $k = 1, 2, 3$ die nichtnegativen Matrizen C und S sowie $D := CS^T$ definiert. Die Matrix D ist links in Abbildung 2.2 dargestellt.

Modelldatensatz 2. Seien y und D wie in Modelldatensatz 1. Durch die Funktionen

$$e_1(\mu) = 0.05(\mu(\mu - 2) + 0.01\mu^3) \quad \text{und}$$
$$e_2(\mu) = 0.06(\exp(-(\mu - 2.5)^2/0.005) - \exp(-(\mu - 2.6)^2/0.005))$$

mit $E'_{i,j} = e_1(y_j)$, $E''_{i,j} = e_2(y_j)$ für $i = 1, \ldots, 101,\ j = 1, \ldots, 401$ seien die Matrizen E' und E'' definiert. Die Matrix E''' enthalte komponentenweise gleichverteilte Zufallswerte im Intervall $[-0.03, 0.03]$. Mit $E := E' + E'' + E'''$ wird die störungsbehaftete Matrix $\tilde{D} := D + E$ definiert. Die Matrizen $\tilde{D}$ sowie E', E'' und E''' sind in Abbildung 2.2 veranschaulicht.

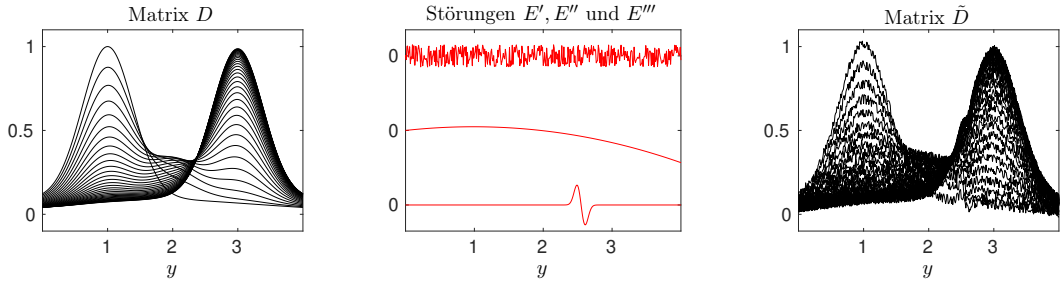


Abbildung 2.2.: Darstellung der Zeilen der nichtnegativen Matrix D (links) aus Modelldatensatz 1, den Störungen E', E'' und E''' (mittig) sowie der Zeilen der Matrix $\tilde{D}$ (rechts) aus Modelldatensatz 2.

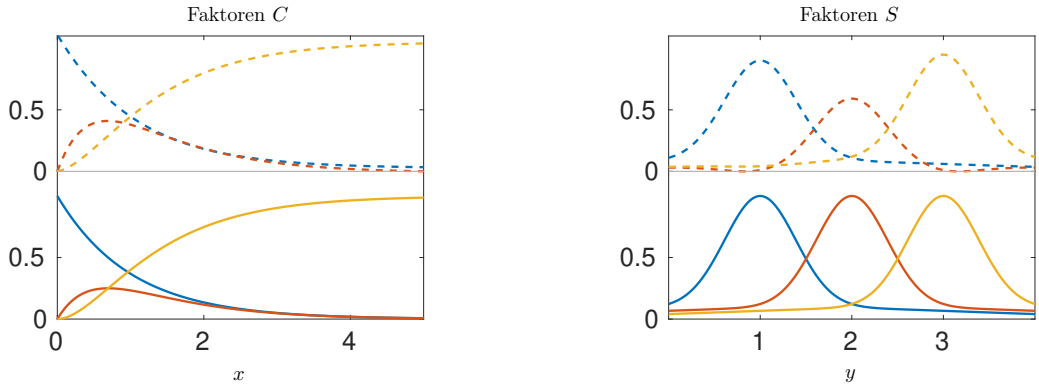


Abbildung 2.3.: Darstellung zweier nichtnegativer Vollrangfaktorisierungen zu D des Modelldatensatzes 1. Die durchgängigen Profile sind den Faktoren zuzuordnen, welche zur Datengenerierung genutzt wurden. Die gestrichelten Profile sind durch Minimierung von (2.5) bestimmt und den entsprechenden Faktoren C^* und S^* des Abschnitts 2.4.3 zuzuordnen.

Für die nichtnegative Matrix D aus Modelldatensatz 1 wird nun die Berechnung einer nichtnegativen Vollrangfaktorisierung veranschaulicht. Die Minimierung von (2.5) wird

mit

$$T^{(0)} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.7}$$

initialisiert. Die resultierenden Faktoren C^* und S^* sind in Abbildung 2.3 dargestellt. Die Werte $\min_{i,j}(C^*_{i,j}) = -6.8 \cdot 10^{-16}$ und $\min_{i,j}(S^*_{i,j}) = -1.2 \cdot 10^{-17}$ liegen im Bereich der Rechengenauigkeit eines Double-Werts und die Faktoren sind als nichtnegativ zu werten.

Für die störungsbehaftete Matrix $\tilde{D}$ aus Modelldatensatz 2 wird nun die Bestimmung von s sowie die Berechnung einer approximativen Matrixfaktorisierung veranschaulicht. Für die in Abbildung 2.4 farblich markierten Singulärwerte ist zu erkennen, dass σ_1, σ_2 und σ_3 signifikant größer als σ_4 sind, womit $s = 3$ geschlussfolgert wird. Damit kann die Niedrigrangapproximation D_s von $\tilde{D}$ betrachtet werden. Wegen $|\min_{i,j}((D_s)_{i,j})| = 0.0058$ wird die Fehlertoleranz $\varepsilon = 0.01$ genutzt und die Minimierung von (2.6) mit $T^{(0)}$ aus (2.7) initialisiert. Die resultierenden Faktoren C^* und S^* sind in Abbildung 2.4 dargestellt und liegen mit $\min_{i,j}(C^*_{i,j}) = -0.01$ und $\min_{i,j}(S^*_{i,j}) = -0.006$ innerhalb der festgelegten Toleranz.

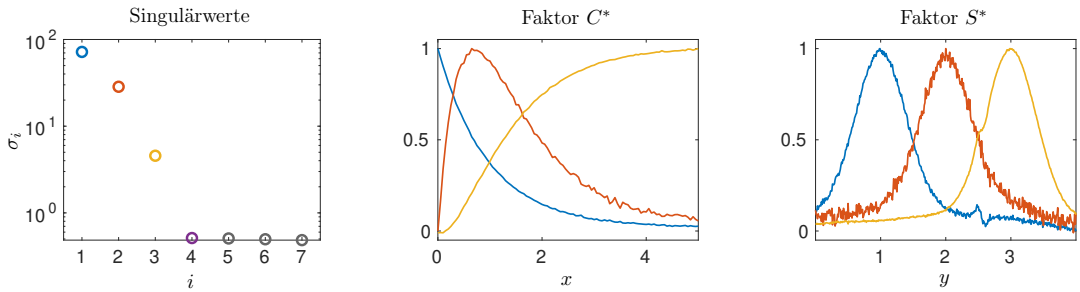


Abbildung 2.4.: Es sind die ersten sieben Singulärwerte von $\tilde{D}$ (links) sowie die numerisch berechneten Faktoren C^* (mittig) und S^* (rechts) dargestellt. Die Spalten beider Faktoren sind so skaliert, dass ihr betragsmäßiges Maximum 1 ist.

2.5. Das regularisierte Matrixfaktorisierungsproblem

Im Fokus dieses Abschnitts steht die regularisierte Matrixfaktorisierung und die Analyse entsprechender Lösungsmengen. Es wird einleitend mit einer Übersicht von Nebenbedingungen begonnen, die zur Regularisierung der Faktorisierungsprobleme 2 und 3 eingesetzt werden können. Dann erfolgt die Beschreibung einer numerischen Methode zur Lösung der regularisierten Matrixfaktorisierung und die Veranschaulichung ausgewählter Ergebnisse basierend auf dem Modelldatensatz 1. Abschließend wird die Regularisierung mittels eines parameterabhängigen Anfangswertproblems betrachtet und daraus die zentrale Fragestellung dieser Arbeit hergeleitet.

Numerische Berechnung

Es wird ein einleitendes Beispiel betrachtet. Dabei ist für eine nichtnegative Matrix $D \in \mathbb{R}^{m \times n}$ mit $\operatorname{rg}(D) = \operatorname{rg}_+(D)$ eine nichtnegative Vollrangfaktorisierung $D = CS^T$ so zu bestimmen, dass für einen festgelegten Index j^* die Ungleichungen $C_{1,j^*} \geq \ldots \geq C_{m,j^*}$

erfüllt sind. Die algorithmische Umsetzung für einen Vektor $v \in \mathbb{R}^m$ erfolgt beispielsweise mittels der Funktion

$$g_{\text{mono}}(v) := \frac{1}{\max_i(|v_i|)} \sum_{i=1}^{m-1} \max(v_{i+1} - v_i, 0)\,. \tag{2.8}$$

Bilden die Elemente von v bezüglich der ansteigenden Indizes eine monoton fallende Folge, gilt $g_{\text{mono}}(v) = 0$ und andernfalls $g_{\text{mono}}(v) > 0$. Eine Erweiterung der Zielfunktion (2.6) ist durch

$$F_{\text{mono}}(T) := F(T) + \delta_1 \cdot g_{\text{mono}}(X_{:,j^*}) \tag{2.9}$$

mit Gewichtungsfaktor δ_1 gegeben.

In Vorbereitung des allgemeinen Falls ist in der Tabelle 2.1 eine Auswahl typischer Nebenbedingungen aufgeführt. Die Funktionen g_i sind darin in Abhängigkeit von T dargestellt. Mit $C = U\Sigma T^+$ und $S = VT^T$ folgt die Konsistenz zur Definition der regularisierten Matrixfaktorisierung in Problem 4.

Name	$g_i(T)$	Effekt
Norm	$\sum_{i=1}^{s} \left\|T_{i,:}V^T\right\|_2$	In Kombination mit der Nichtnegativität von S erfolgt eine Verringerung des Integrals der Profile
Glattheit	$\sum_{i=1}^{s} \left\|\frac{\Delta^2}{\Delta x^2} T_{i,:}V^T\right\|_2$	Bevorzugung glatter Profile
Vorgabe eines eventuell parameterabhängigen Modells	$\left\|S^{\text{mod}}(p) - TV^T\right\|_F$	Konsistenz mit einem vorgegebenen Modell; Bestimmung optimaler Parameter p^*
Equality constraint; Fixieren von Spalten in den Faktoren (exemplarisch durch v für die i-te Spalte von S) [41, 122]	$\left\|v - S(:,i)^T\right\|_2 = \left\|v - T_{i,:}V^T\right\|_2$	Übereinstimmung mit separat zugänglichen Profilen
Monotonie, Unimodalität [120]	siehe Gleichung (2.8) und [109]	Bevorzugung monoton fallender, steigender oder unimodaler Profile
Closure constraint; Einhalten von Bilanzgleichungen (wie etwa die Massenbilanz) [120]	$\left\|(\sum_{i=1}^{s} T_{i,:})V^T - \alpha(1,\ldots,1)\right\|_2$ mit $\alpha > 0$	Bevorzugung von Faktoren mit konstanten Zeilensummen
Selectivity constraint; Festlegen eines Indexbereichs I mit nur einem aktiven Profil	$\sum_{i=1, i\neq j}^{s} \left\|T_{i,:}V_{:,I}^T\right\|_2$	Unterdrücken von $s-1$ Profilen im Indexbereich I
Local rank; Festlegen eines Indexbereichs I für den nur eine vorher definierte Anzahl p an Profilen von Null verschiedene Einträge hat	$\left\lvert \text{rg}\left(TV_{:,I}^T\right) - p\right\rvert$	Verallgemeinerung der *selectivity constraint* für eine p-elementige Menge an Profilen

Tabelle 2.1.: Übersicht möglicher Nebenbedingungen zur Verwendung bei der Minimierung der Zielfunktion (2.10). Die genannten Nebenbedingungen sind analog auch für den Faktor C definiert. Weitere Möglichkeiten zur Regularisierung sind in [17, 60, 119] zu finden.

Eine Verallgemeinerung der Zielfunktion (2.9) des einleitenden Beispiels für eine Auswahl

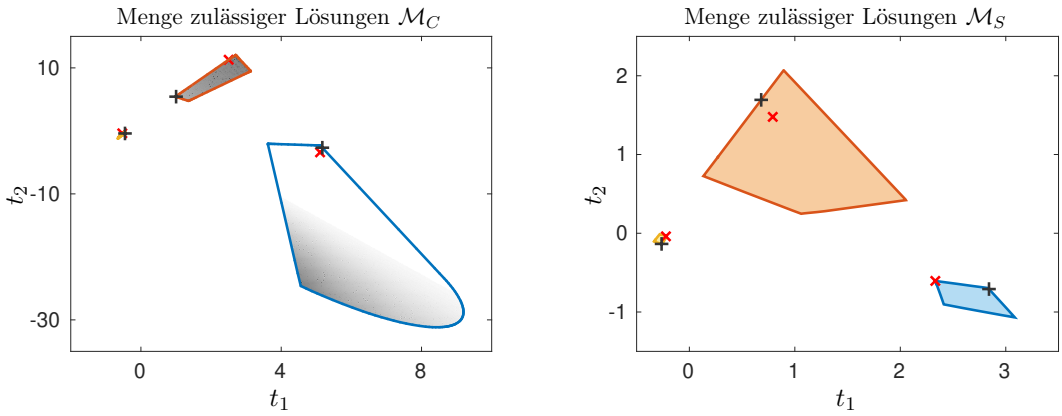


Abbildung 2.5.: Veranschaulichung des Einflusses der in (2.9) genutzten Minimierung zur Bestimmung einer nichtnegativen Vollrangfaktorisierung für D aus Modelldatensatz 1. Die schwarzen Markierungen repräsentieren die Faktoren C^* und S^* als Resultat der Minimierung von (2.9). Die roten Markierungen entsprechen den Faktoren, die zur Datengenerierung genutzt wurden. Die Schattierung in $\mathcal{M}_C$ veranschaulicht die Übereinstimmung eines Punktes mit der Nebenbedingung, wobei helle Bereiche "monoton fallenden" Profilen zugeordnet werden.

von r verschiedenen Nebenbedingungen $g_1(T), \ldots, g_r(T)$ mit Gewichten $\delta_1, \ldots, \delta_r$ ist

$$F_{\text{reg}}(T) := F(T) + \sum_{i=1}^{r} \delta_i g_i(T) \ . \tag{2.10}$$

Die Gewichte δ_i werden üblicherweise um mindestens eine Größenordnung kleiner als die Gewichte γ_i in $F(T)$ gewählt, da sich die Nebenbedingungen $g_i(T)$ möglichst nicht auf die Einhaltung der Nichtnegativität der Faktoren oder der Regularität der Matrix T auswirken sollen.

Abschließend wird auf eine Reihe von Programmen zur Bestimmung von regularisierten Matrixfaktorisierungen hingewiesen. *MCR-ALS* [63] ist eine sehr populäre und umfangreiche Matlab-Toolbox. Außerdem sind das Excel/Matlab-basierte *ReactLab* [80], die *Peakgruppenanalyse* [104], die *Band-target entropy minimization* [18] und *Peaxact* aus der *S-Pact* Toolbox [4] zu nennen.

Die regularisierte Matrixfaktorisierung für ein Modellproblem

Exemplarisch wird nun die Auswirkung der Nebenbedingung g_{mono} aus (2.9) mit $j^* = 1$ anhand des Modelldatensatzes 1 demonstriert. Hierzu sind in Abbildung 2.5 die Mengen zulässiger Lösungen für die Faktoren C und S dargestellt. Zusätzlich erfolgt die Auswertung der Nebenbedingung innerhalb der Segmente von $\mathcal{M}_C$. Die graue Schattierung in $\mathcal{M}_C$ korreliert mit dem Funktionswert von $g_{\text{mono}}(U\Sigma(t_1, t_2))$, wobei weiße Bereiche dem Wert 0 entsprechen. Es sei insbesondere auf das blaue Segment von $\mathcal{M}_C$ hingewiesen, welches der ersten Spalte von X zugeordnet werden kann. Durch Minimierung von (2.9) ausgehend von $T^{(0)}$ aus (2.7) wurde zusätzlich eine Faktorisierung $D = C^*(S^*)^T$ berechnet. Die drei schwarzen Markierungen in $\mathcal{M}_C$ repräsentieren C^* und in $\mathcal{M}_S$ den Faktor S^*. Sie liegen innerhalb der Grenzen der jeweiligen Segmente von $\mathcal{M}_C$ und $\mathcal{M}_S$, womit C^* und S^* nichtnegativ sind. Weiter liegt die schwarze Markierung im blauen Segment im weißen Bereich, womit auch die zusätzliche Nebenbedingung eingehalten ist.

Ergänzend sind in Abbildung 2.6 weitere Faktorisierungen gezeigt, die durch Minimierung

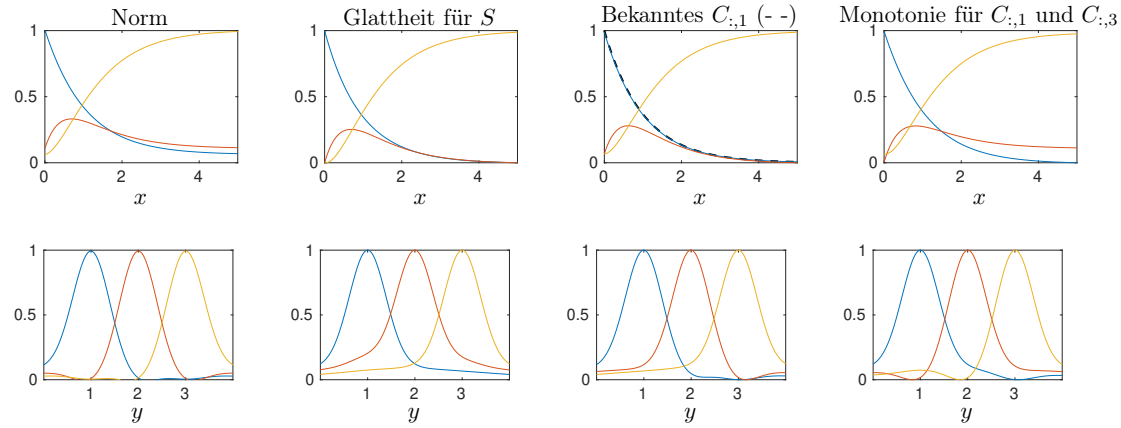


Abbildung 2.6.: Es sind sind Faktoren C^* (oben) und S^*(unten) dargestellt, die durch die Minimierung von (2.10) für das Modellproblem 1 ermittelt wurden. Zur Berechnung, der in den Spalten der Abbildung gezeigten Faktoren, wurde die im jeweiligen Titel genannte Nebenbedingung genutzt.

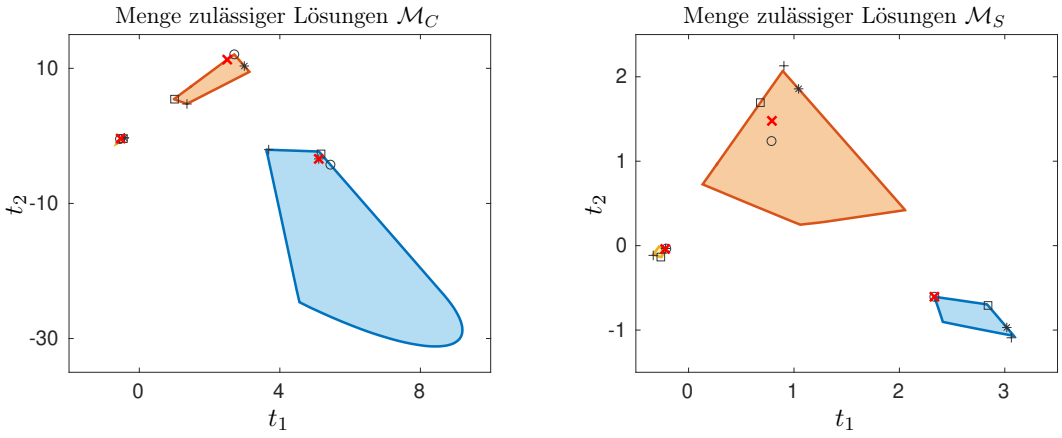


Abbildung 2.7.: Es sind die Projektionen, der in Abbildung 2.6 gezeigten Faktorisierungen, in die Mengen $\mathcal{M}_C$ und $\mathcal{M}_S$ dargestellt: Norm (+), Glattheit (∘), bekanntes erstes Profil in C (∗) und "Monotonie" des ersten und dritten Profils in C (□). Die zur Datengenerierung genutzten Faktoren C und S werden durch die roten Kreuze repräsentiert.

von (2.10) mit den jeweiligen Nebenbedingungen berechnet wurden. Ihre Projektionen in die Mengen zulässiger Lösungen sind in Abbildung 2.7 dargestellt. Für den Modelldatensatz 1 ist die Verwendung der Glattheits-Nebenbedingung besonders geeignet. Die berechneten Faktoren weichen nur gering von den zur Datengenerierung genutzten Faktoren C und S ab. Sei hierzu exemplarisch der relative Fehler

$$\|S - S^{\text{Glatt}}\|_F / \|S\|_F = 0.018$$

genannt. Auch die Kombination mehrerer Funktionen $g_i(T)$ und Variationen der Gewichtung können eingesetzt werden, um bestmögliche Ergebnisse zu erzielen. Ist im Anwendungskontext nur genau eine zulässige Faktorisierung sinnvoll interpretierbar, ist eine Regularisierung des jeweiligen Faktorisierungsproblems typischerweise unerlässlich.

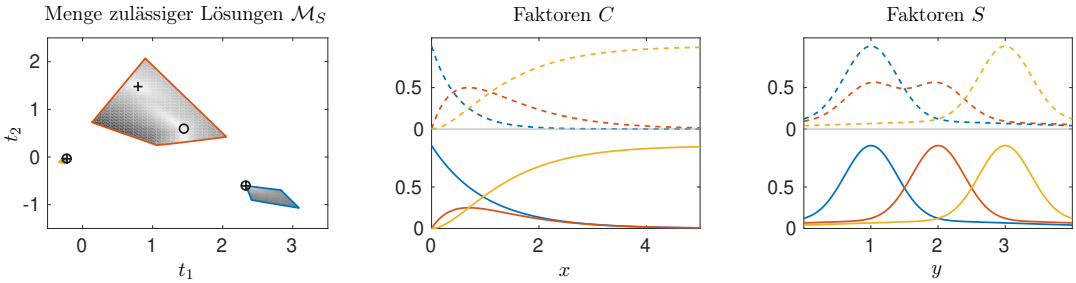


Abbildung 2.8.: Links ist die Menge zulässiger Lösungen $\mathcal{M}_S$ für den Modelldatensatz 1 zu sehen. Es sind zwei verschiedene nichtnegative Vollrangfaktorisierungen durch Kreuze beziehungsweise Kreise markiert. Beide Faktoren C sind konsistent mit dem Anfangswertproblem. Die Schattierung korreliert mit der Abweichung von dieser Nebenbedingung. Mittig sind die Profile der Faktoren C^* und rechts die Profile der Faktoren S^* der zwei Faktorisierungen dargestellt.

Regularisierung durch parameterabhängige Anfangswertprobleme

Das Modellproblem 1 des vorherigen Abschnitts steht erneut im Fokus. Es wird nun eine Nebenbedingung genutzt, die die Konsistenz des Faktors C mit dem parameterabhängigen Anfangswertproblem

$$\begin{pmatrix}\dot{c}_1(\xi)\\ \dot{c}_2(\xi)\\ \dot{c}_3(\xi)\end{pmatrix} = \begin{pmatrix}-k_1 & 0 & 0\\ k_1 & -k_2 & 0\\ 0 & k_2 & 0\end{pmatrix}\begin{pmatrix}c_1(\xi)\\ c_2(\xi)\\ c_3(\xi)\end{pmatrix} \quad \text{mit} \quad \begin{pmatrix}c_1(0)\\ c_2(0)\\ c_3(0)\end{pmatrix} = \begin{pmatrix}1\\ 0\\ 0\end{pmatrix} \quad \text{und} \quad k \in \mathbb{R}^2_+ \tag{2.11}$$

fordert. Es wird zunächst der Konsistenzbegriff erläutert. Sei hierzu $c : \mathbb{R}_{\geq 0} \to \mathbb{R}^3_{\geq 0}$ die Lösung des Anfangswertproblems zum Parameter k. Existiert ein $k \in \mathbb{R}^2_+$, sodass mit dem Gitter $x = (0,\ 0.05, \ldots, 5)^T$ die Gleichungen $C_{i,:} = c(x_i)^T,\ i = 1, \ldots, 101$ erfüllt sind, heißt C konsistent zum Anfangswertproblem. In Abbildung 2.8 ist analog zu Abbildung 2.5 die Auswertung der Menge zulässiger Lösungen $\mathcal{M}_S$ bezüglich der beschriebenen Nebenbedingung dargestellt. Für die Auswertung eines Punktes $(t_1, t_2) \in \mathcal{M}_S$ wird geprüft, ob sich das entsprechende Profil $(1, t_1, t_2)V^T$ zu einer nichtnegativen Vollrangfaktorisierung ergänzen lässt, deren Faktor C zudem konsistent mit dem Anfangswertproblem (2.11) ist. Der Abbildung ist zu entnehmen, dass zwei zu (2.11) konsistente nichtnegative Vollrangfaktorisierungen existieren. Unter Verwendung dieser Nebenbedingung kann also keine eindeutige Lösung des regularisierten Matrixfaktorisierungsproblems bestimmt werden.

Aus dieser Beobachtung resultieren die folgenden Fragestellungen:

1. Für welche Anfangswertprobleme hat die regularisierte Matrixfaktorisierung genau eine Lösung?
2. Wenn eine nichttriviale Lösungsmenge existiert, lässt sie sich mittels expliziter oder impliziter Gleichungen beschreiben?
3. Können Lösungen solcher regularisierter Matrixfaktorisierungen eindeutig im Raum der Parameter k dargestellt werden?
4. Lassen sich entsprechende Lösungsmengen numerisch approximieren?

Diese Fragen werden in den folgenden Kapiteln detailliert diskutiert.

2.6. Anwendungen der nichtnegativen Matrixfaktorisierung

Abschließend wird auf Anwendungen der nichtnegativen Matrixfaktorisierung sowie ihrer Modifikationen eingegangen. Es muss hierbei unterschieden werden, ob das Ziel in der Bestimmung einer beliebigen nichtnegativen Faktorisierung oder einer Faktorisierung mit weiteren speziellen Eigenschaften besteht. Der erste Fall ist der einfachere, da nur ein Element der Lösungsmenge bestimmt werden muss. Die entsprechenden Anwendungen sind dennoch vielfältig. In [76] werden durch die Berechnung einer nichtnegativen Matrixfaktorisierung Bilder von Gesichtern in ihre signifikanten Bereiche zerlegt, die eine anschließende Klassifizierung ermöglichen. Die Arbeit [52] beschreibt einen auf der nichtnegativen Matrixfaktorisierung basierenden Indikator zur Entdeckung von Eindringlingen in Computernetzwerken. Eine weitere, in den Bereich der Klassifizierung einzuordnende, Veröffentlichung [132] nutzt die nichtnegative Matrixfaktorisierung als Datenvorbehandlungsschritt um signifikante Gene und Genbereiche zur Tumorerkennung zu identifizieren. In [10] wird mittels nichtnegativer Matrixfaktorisierung eine inhaltliche Gruppierung von großen Mengen an Emails vorgenommen. Die zweite Klasse von Fragestellungen zielt auf die Bestimmung einer ganz speziellen (unbekannten) Faktorisierung innerhalb der Lösungsmenge des jeweiligen Faktorisierungsproblems ab. Nur diese eine Faktorisierung spiegelt den problembezogenen Sachverhalt korrekt wider. In [23] werden mittels regularisierter Matrixfaktorisierung Börsendaten analysiert, um Trends in deren zeitaufgelösten Preisverläufen zu extrahieren. Weitere Anwendungen sind in [14, 20, 91] zu finden.

Im Folgenden wird näher auf die nichtnegative Vollrangfaktorisierung im Kontext chemometrischer Anwendungen eingegangen.

Die nichtnegative Matrixfaktorisierung in der Chemometrie

Spektroskopische Messmethoden bieten Möglichkeiten chemische Reaktionssysteme oder Prozesse sowie die daran beteiligten Spezies zu analysieren [50, 71, 111, 117]. Durch Ausnutzen messbarer Eigenschaften, wie zum Beispiel der Wechselwirkung elektromagnetischer Strahlung im sichtbaren und ultravioletten Bereich (UV/Vis-Spektroskopie) oder im infraroten Bereich (FTIR-Spektroskopie, [59]), lassen sich Spektren bestimmen, die beispielsweise die Identifizierung unbekannter Spezies in einem Reaktionssystem ermöglichen.

Sei hierzu $d \in \mathbb{R}^n$ ein solches nichtnegatives Spektrum eines Stoffgemisches aus s verschiedenen Spezies $\mathcal{X}_1, \ldots, \mathcal{X}_s$ auf einem Frequenzgitter. Die für diese Arbeit interessanten Messmethoden haben gemein, dass den verschiedenen Spezies nichtnegative *Reinspektren* $d_{\mathcal{X}_1}, \ldots, d_{\mathcal{X}_s} \in \mathbb{R}^n$ zugeordnet werden können, deren Skalierung linear von der jeweiligen Konzentration $c_{\mathcal{X}_1}, \ldots, c_{\mathcal{X}_s} \in \mathbb{R}$ innerhalb des Stoffgemisches abhängt. Aus dem Gesetz von Lambert-Beer [9] folgt, dass sich das Spektrum d als Linearkombination

$$d = \sum_{i=1}^{s} c_{\mathcal{X}_i} \cdot d_{\mathcal{X}_i}^T \tag{2.12}$$

der Reinspektren darstellen lässt. Die Konzentrationen $c_{\mathcal{X}_1}, \ldots, c_{\mathcal{X}_s}$ und somit auch das Spektrum d ändern sich typischerweise mit fortlaufender Reaktion. Eine zeitabhängige

Darstellung der Gleichung (2.12) ist durch

$$d(t) = \sum_{i=1}^{s} c_{\mathcal{X}_i}(t) \cdot d_{\mathcal{X}_i} = \begin{pmatrix} c_{\mathcal{X}_1}(t) & \dots & c_{\mathcal{X}_s}(t) \end{pmatrix} \cdot \underbrace{\begin{pmatrix} d_{\mathcal{X}_1}^T \\ \vdots \\ d_{\mathcal{X}_s}^T \end{pmatrix}}_{S^T}$$

mit dem (nichtnegativen) *Spektrenfaktor* $S \in \mathbb{R}^{n \times s}$ gegeben. Weil die Messung der Spektren nicht kontinuierlich, sondern auf einem zeitdiskreten Gitter $T_G = (t_1, \dots, t_m)$ durchgeführt wird, ist $d(t)$ nur für $t \in T_G$ bekannt. Damit folgt

$$\underbrace{\begin{pmatrix} d(t_1) \\ \vdots \\ d(t_m) \end{pmatrix}}_{D \in \mathbb{R}^{m \times n}} = \underbrace{\begin{pmatrix} c_{\mathcal{X}_1}(t_1) & \cdots & c_{\mathcal{X}_s}(t_1) \\ \vdots & & \vdots \\ c_{\mathcal{X}_1}(t_m) & \cdots & c_{\mathcal{X}_s}(t_m) \end{pmatrix}}_{C \in \mathbb{R}^{m \times s}} S^T \;\Leftrightarrow\; D = CS^T,$$

wobei D zeilenweise die Spektren $d(t_i)$ für $i = 1, \dots, m$ und der *Konzentrationsfaktor* C spaltenweise die zeitdiskreten Konzentrationsverläufe der Spezies $\mathcal{X}_1, \dots, \mathcal{X}_s$ enthält. Es sei angemerkt, dass abseits von Spezialfällen lediglich D messbar und somit gegeben ist. Die weiteren Größen s, C, S sind im Regelfall unbekannt. Die Bestimmung von s erfolgt typischerweise auf Basis zusätzlicher Informationen über das chemische Reaktionssystem oder kann, wie in Abschnitt 2.4.2 beschrieben, mittels der Singulärwerte approximiert werden. Weil für den beschriebenen chemischen Sachverhalt unter Berücksichtigung von Störeinflüssen die Zeilen von D Superpositionen von nichtnegativen Reinspektren sind, kann davon ausgegangen werden, dass für ein korrekt gewähltes s eine approximative Matrixfaktorisierung existiert. Der Fokus liegt folglich auf der geeigneten Bestimmung der verbleibenden unbekannten Faktoren C und S. Die Aufgabe der Bestimmung von s sowie der Berechnung der Faktoren C und S wird im Kontext der Chemometrie unter dem Begriff der *Reinkomponentenzerlegung* zusammengefasst.

Die Anwendungsvarianten sind vielfältig und ohne den Anspruch auf Vollständigkeit seien die "klassische" Faktorisierung einer einzelnen Matrix [7,43], die simultane Analyse einer Menge von mindestens zwei Matrizen (englisch: multisets) [25, 30] und die Auswertung von hyperspektralen Bildern [61,131] genannt. Einen sehr guten Einstieg in das Themengebiet der approximativen Matrixfaktorisierung im Sinne der Reinkomponentenzerlegung bieten [81, 82].

2.7. Kritische Zusammenfassung

Im Fokus dieses Kapitels stehen die nichtnegative Vollrangfaktorisierung für idealisierte Ausgangsdaten, die approximative Matrixfaktorisierung zur Berücksichtigung von Störungen und das Problem der regularisierten Matrixfaktorisierung, welches zusätzlich das Einbinden von Nebenbedingungen beinhaltet. Es werden numerische Lösungsmethoden für die Faktorisierungsaufgaben beschrieben und mit der Menge zulässiger Lösungen eine Möglichkeit zur grafischen Darstellung der Lösungsmengen dieser Probleme erläutert. Es wird festgestellt, dass in allen betrachteten Fällen nichttrivialen Lösungsmengen auftreten können.

Die folgenden Punkte sind kritisch anzumerken:

- Die Ergebnisse dieses Kapitels stellen die Grundlage für die weitere Arbeit dar. In Kapitel 3 erfolgt eine detaillierte Betrachtung von Lösungsmethoden für die bereits kurz in Abschnitt 2.5 behandelte regularisierte Matrixfaktorisierung unter Verwendung parameterabhängiger Anfangswertprobleme. In Kapitel 4 wird die Idee der niedrigdimensionalen Darstellung von nichtnegativen Matrixfaktorisierungen erweitert. Hierzu wird eine Darstellung von Lösungen der regularisierten Matrixfaktorisierung im Raum der Parameter des entsprechenden Anfangswertproblems eingeführt und analysiert.
- Die Lösungsmengen der genannten Faktorisierungsprobleme lassen sich zum Beispiel mit der Menge zulässiger Lösungen darstellen. Ihr Nutzen zur Lösung von Anwendungsproblemen wird durch die notwendige grafische Darstellung auf Fälle mit $s \leq 4$ limitiert. Jede Möglichkeit zur Reduktion der Komponentenanzahl s eines Anwendungsproblems, wie beispielsweise die Betrachtung von Subsystemen, sollte genutzt werden. Dieser Gedanke überträgt sich auch auf die später in dieser Arbeit betrachteten Parameterlösungsmengen von Anfangswertproblemen.
- Für anwendungsrelevante Probleme kann stets vom Vorhandensein von Störeinflüssen ausgegangen werden, womit die regularisierte Matrixfaktorisierung von besonderer Wichtigkeit ist. Im Gegensatz dazu lassen sich theoretische Betrachtungen und die damit verbundene Methodenentwicklung oft nur für idealisierte Ausgangsdaten umsetzen. Diese Herangehensweisen sollten sinnvoll miteinander kombiniert werden. Methoden zur Analyse störungsfreier Daten lassen sich etwa durch die Akzeptanz kleiner Fehler auch für gestörte Ausgangsdaten einsetzen und erzielen erfahrungsgemäß gute Ergebnisse.

3. Kinetische Modellierung

In diesem Kapitel werden sogenannte kinetische Modelle und deren Verwendung als Nebenbedingung in der regularisierten Matrixfaktorisierungsaufgabe 4 betrachtet. Wegen des starken chemometrischen Bezugs werden die in Abschnitt 2.6 eingeführten Begriffe genutzt[1]. Ein kinetisches Modell ist dabei nichts anderes als ein parameterabhängiges Anfangswertproblem, das sich für die Modellierung des Konzentrationsfaktors C eignet. Diese Modelle lassen sich typischerweise aus der Struktur des zugrunde liegenden Reaktionssystems herleiten [83]. Die Parameter des Modells können simultan zu den Faktoren C und S ermittelt werden. Trotz einer gegebenenfalls höheren Anzahl an Freiheitsgraden des regularisierten Problems gegenüber dem unregularisierten, stellt die Kenntnis eines kinetischen Modells bei der Bestimmung einer regularisierten Matrixfaktorisierung eine wertvolle Information dar. Zudem wird oft eine Stabilisierung der numerischen Lösungsmethoden beobachtet.

In Abschnitt 3.1 werden zwei Varianten zur Verwendung von kinetischen Modellen im Kontext des Problems der nichtnegativen Matrixfaktorisierung vorgestellt [24,26,48]. Solche Modelle müssen nicht zwangsweise alle Spalten des Konzentrationsfaktors beschreiben und umgekehrt müssen zu einer Folge aus Spektren nicht immer alle vom Modell beschriebenen Komponenten beitragen [71]. In Abschnitt 3.2 werden hierzu entsprechende Modifikationen der bisher eingeführten Algorithmen präsentiert. Eine Aussage über die Nichtnegativität der betrachteten Modelle wird in Abschnitt 3.3 bewiesen. Dass die Nutzung eines kinetischen Modells nicht zwangsläufig auf eine eindeutige Faktorisierung führt, wird in Abschnitt 3.4 anhand zweier Beispiele gezeigt. Gleichzeitig erfolgt damit die Motivation des zentralen Kapitels 4 dieser Arbeit.

3.1. Kinetiken als Soft- und Hard-Modelle

Einleitend wird in diesem Abschnitt erläutert, wie mathematische Modelle in Form von parameterabhängigen Systemen gewöhnlicher Differentialgleichungen zur Modellierung von Konzentrationsverläufen chemischer Reaktionssysteme genutzt werden können. Anschließend werden zwei Ansätze präsentiert, die das Einbinden solcher Modelle in die Zielfunktion (2.6) aus Abschnitt 2.4.2 ermöglichen.

Herleitung kinetischer Modelle aus Reaktionsgleichungen

Die Modellierung von Konzentrationsprofilen durch Anfangswertprobleme ist in der Literatur ausführlich beschrieben [5, 83]. Zur Einführung wird der einfache Fall einer uni-

[1] Die *Reinkomponentenzerlegung* einer spektroskopischen Messfolge D umfasst die Bestimmung der Anzahl an Komponenten s, des *Konzentrationsfaktors* C und des *Spektrenfaktors* S durch Lösung einer nichtnegativen Vollrangfaktorisierung $D = CS^T$.

molekularen Reaktion mit der Reaktionsgleichung

$$\mathcal{X} \xrightarrow{k_1} \mathcal{Y} \tag{3.1}$$

betrachtet. Die zeitabhängigen Konzentrationsverläufe der Komponenten $\mathcal{X}$ und $\mathcal{Y}$ seien durch die Funktionen $c_\mathcal{X}(t), c_\mathcal{Y}(t) : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ beschrieben. Die entsprechenden Ableitungen $\dot{c}_\mathcal{X}(t), \dot{c}_\mathcal{Y}(t)$ hängen von $c_\mathcal{X}(t)$ und $k_1 \in \mathbb{R}_+$, einem sogenannten Geschwindigkeitsparameter[2], ab und lassen sich als gewöhnliches Differentialgleichungssystem

$$\begin{pmatrix} \dot{c}_\mathcal{X}(t) \\ \dot{c}_\mathcal{Y}(t) \end{pmatrix} = \begin{pmatrix} -k_1 & 0 \\ k_1 & 0 \end{pmatrix} \begin{pmatrix} c_\mathcal{X}(t) \\ c_\mathcal{Y}(t) \end{pmatrix} \tag{3.2}$$

formulieren.

Davon ausgehend werden zwei Verallgemeinerungen betrachtet, um beliebige Reaktionsmodelle aufstellen zu können. Als erstes wird das Auftreten simultan ablaufender Reaktionen innerhalb eines Systems behandelt. Die rechten Seiten der jeweiligen Differentialgleichungssysteme werden dabei addiert. Unter Berücksichtigung einer Rückreaktion $\mathcal{Y} \to \mathcal{X}$ mit Geschwindigkeitsparameter $k_2 \in \mathbb{R}_+$ in (3.1) resultiert mit (3.2)

$$\underbrace{\begin{pmatrix} \dot{c}_\mathcal{X}(t) \\ \dot{c}_\mathcal{Y}(t) \end{pmatrix} = \begin{pmatrix} 0 & k_2 \\ 0 & -k_2 \end{pmatrix} \begin{pmatrix} c_\mathcal{X}(t) \\ c_\mathcal{Y}(t) \end{pmatrix}}_{\text{Rückreaktion}} \quad \rightsquigarrow \quad \underbrace{\begin{pmatrix} \dot{c}_\mathcal{X}(t) \\ \dot{c}_\mathcal{Y}(t) \end{pmatrix} = \begin{pmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix} \begin{pmatrix} c_\mathcal{X}(t) \\ c_\mathcal{Y}(t) \end{pmatrix}}_{\text{simultane Betrachtung}} .$$

Als zweiter Schritt werden Reaktionsgleichungen der allgemeinen Form

$$\alpha_1\mathcal{X}_1 + \ldots + \alpha_e\mathcal{X}_e \xrightarrow{k_1} \beta_1\mathcal{Y}_1 + \ldots + \beta_p\mathcal{Y}_p$$

mit e Reaktanten $\mathcal{X}_1, \ldots, \mathcal{X}_e$ und p Produkten $\mathcal{Y}_1, \ldots, \mathcal{Y}_p$ betrachtet. Das zugehörige Differentialgleichungssystem lautet dann

$$\begin{pmatrix} \dot{c}_{\mathcal{X}_1}(t) \\ \vdots \\ \dot{c}_{\mathcal{X}_e}(t) \\ \dot{c}_{\mathcal{Y}_1}(t) \\ \vdots \\ \dot{c}_{\mathcal{Y}_p}(t) \end{pmatrix} = \begin{pmatrix} -\alpha_1 k_1 \\ \vdots \\ -\alpha_e k_1 \\ \beta_1 k_1 \\ \vdots \\ \beta_p k_1 \end{pmatrix} (c_{\mathcal{X}_1}(t))^{\alpha_1} \cdot \ldots \cdot (c_{\mathcal{X}_e}(t))^{\alpha_e}$$

mit $\alpha \in \mathbb{N}^e$, $\beta \in \mathbb{N}^p$ und $k_1 \in \mathbb{R}_+$. Der Wert $\sum_i \alpha_i$ wird *Reaktionsordnung* genannt. Es sei angemerkt, dass die in dieser Arbeit relevanten Fälle typischerweise eine maximale Reaktionsordnung von 2 haben[3]. Durch Kombination der zwei Schritte ist es möglich, allgemeine Reaktionsmodelle in die Form eines Differentialgleichungssystems zu überführen.

[2] Geschwindigkeitsparameter werden auch häufig als *Geschwindigkeitskonstanten* bezeichnet.

[3] Bereits eine Reaktionsordnung von 3 besagt, dass für den Ablauf dieser Reaktion drei Moleküle zusammentreffen müssen. Die Wahrscheinlichkeit hierfür gegenüber einer Verkettung von Reaktion mit geringerer Ordnung ist vernachlässigbar klein.

Beispiel 1. *Das Differentialgleichungssystem zu einem Reaktionssystem*

$$2\mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \xrightarrow{k_3} \mathcal{Z}$$

lautet

$$\begin{pmatrix} \dot{c}_{\mathcal{X}}(t) \\ \dot{c}_{\mathcal{Y}}(t) \\ \dot{c}_{\mathcal{Z}}(t) \end{pmatrix} = \begin{pmatrix} -2k_1 & 2k_2 \\ k_1 & -k_2-k_3 \\ 0 & k_3 \end{pmatrix} \begin{pmatrix} c_{\mathcal{X}}(t)^2 \\ c_{\mathcal{Y}}(t) \end{pmatrix}.$$

Im allgemeinen Fall lässt sich ein Reaktionsmodell aus q Teilreaktionen unter Kenntnis von Anfangswerten (respektive Anfangskonzentrationen) c_0 zum Zeitpunkt t_0 durch ein Anfangswertproblem der Form

$$\dot{c}(t) = M(k)f(c(t)), \quad c(t_0) = c_0 \geq 0 \tag{3.3}$$

beschreiben. Die Matrix $M(k)$ hängt nur vom Parametervektor $k = (k_1, \ldots, k_q)^T \in \mathbb{R}_+^q$ ab und hat immer s Zeilen. Die Anzahl der Spalten variiert je nach betrachtetem Reaktionsmodell. Die Funktion f bildet die vektorwertige Funktion $c(t)$ auf einen Vektor relevanter Produkte der Komponenten von $c(t)$ ab (siehe Beispiel 1 mit $f(x) = (x_1^2, x_2)^T$). Ein Anfangswertproblem wie in (3.3), welches sich aus einem Reaktionssystem ableitet, wird im Folgenden als *kinetisches Modell* oder kurz *Kinetik* bezeichnet. Die Ordnung eines kinetischen Modells entspricht der maximalen Reaktionsordnung unter allen daran beteiligten Teilreaktionen. Da das Differentialgleichungssystem in (3.3) autonom ist, liegt Invarianz der Lösung $c(t)$ bezüglich der Variablen t vor. Folglich kann durch eine geeignete Substitution $c(t) \to c(t - t_0)$ davon ausgegangen werden, dass ohne Beschränkung der Allgemeinheit $t_0 = 0$ gilt. Kinetische Modelle erster Ordnung spielen im weiteren Verlauf dieser Arbeit eine besondere Rolle. Für sie gilt $f(c(t)) = c(t)$ und $M(k)$ ist eine $s \times s$ Matrix.

Um ein kinetisches Modell zur Regularisierung der Faktorisierungsprobleme aus Kapitel 2 zu nutzen, ist es erforderlich, dieses auf einem diskreten Gitter $(t_1, \ldots, t_m)$ auszuwerten[4]. Die resultierende Matrix lautet

$$C^{\mathrm{dgl}}(k) = \begin{pmatrix} c_1(t_1) & \cdots & c_s(t_1) \\ \vdots & & \vdots \\ c_1(t_m) & \cdots & c_s(t_m) \end{pmatrix} \in \mathbb{R}^{m \times s}.$$

Zur Bestimmung von $C^{\mathrm{dgl}}(k)$ ist also das Lösen des von k abhängigen Anfangswertproblems (3.3) notwendig. Wenn es die Komplexität des Differentialgleichungssystems zulässt, kann eine analytische Lösung bestimmt und eine Auswertung auf dem Zeitgitter vorgenommen werden. Üblicherweise ist eine solche Lösung nicht zugänglich und es muss auf numerische Lösungsverfahren, wie beispielsweise die Klasse der Runge-Kutta-Verfahren [32, 37, 54], zurückgegriffen werden.

Numerische Lösung der Reinkomponentenzerlegung mittels kinetischer Modelle

Es werden nun zwei Ansätze zur Lösung der durch ein kinetisches Modell regularisierten Matrixfaktorisierung beschrieben. Es wird mit einer Adaption des Ansatzes aus Abschnitt

[4] Für gewöhnlich ergibt sich das Gitter aus dem jeweiligen Anwendungsproblem.

2.4.2 begonnen. Weil kinetische Modelle von Geschwindigkeitsparametern k abhängen, müssen diese ebenfalls in der zugrunde liegenden Minimierung berücksichtigt werden. Es wird also simultan zur Matrix T auch für den Parameter k ein optimaler Wert k^* bestimmt. Die Zielfunktion (2.6) wird entsprechend angepasst und lautet

$$F_{\text{soft}}(T,k) := F(T) + \delta_1 \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{s}\left(\frac{C_{i,j} - (C^{\text{dgl}}(k))_{i,j}}{\max_l(C_{l,j})}\right)^2}_{=:F_{\text{dgl}}(C,k)=F_{\text{dgl}}} \tag{3.4}$$

mit $C := U\Sigma T^+$ und Gewichtungsfaktor δ_1. Mit δ_1 kann die Relevanz der Übereinstimmung von Modell und Konzentrationsfaktor für die Minimierung festgelegt werden. Typischerweise wird δ_1 gegenüber den Gewichtungen γ_1, γ_2 und γ_3 in $F(T)$ um mindestens eine Größenordnung kleiner gewählt. In diesem Fall wird häufig der Begriff *Soft*-Modell genutzt. Die Konsistenz der Konzentrationsverläufe C mit dem kinetischen Modell ist dann erwünscht, aber nicht zwingend erforderlich. Der Begriff "Konsistenz mit einer Kinetik" ist so zu verstehen, dass zu einer Faktorisierung von D mit Konzentrationsfaktor C ein Parameter k des kinetischen Modells existiert, sodass $C = C^{\text{dgl}}(k)$ gilt.

Wird die Konsistenz ohnehin vorausgesetzt, kann auch eine Implementierungen als sogenanntes *Hard*-Modell verwendet werden. Diese Variante ist häufig numerisch stabiler und effizienter, setzt aber auch voraus, dass das gewählte kinetische Modell korrekt ist. Folgende Überlegungen sind für diesen Ansatz wichtig: Es wird angenommen, dass die korrekte Kinetik ausgewählt ist und eine Matrix T sowie ein Geschwindigkeitsparameter k existieren, sodass

$$C^{\text{dgl}}(k) = U\Sigma T^+,\ T^+T - I = 0,\ U\Sigma T^+ \geq 0 \text{ und } TV^T \geq 0$$

gelten. Die Idee des Hard-Modell-Ansatzes besteht darin, abhängig vom Parametervektor k eine optimale Transformationsmatrix T zu bestimmen, die

$$\|C^{\text{dgl}}(k) - U\Sigma T^+\|_F \to \min \tag{3.5}$$

löst, womit eine Zielfunktion definiert werden kann, die nur von k abhängt. Nach [49] ist $T = ((U\Sigma)^+ C^{\text{dgl}}(k))^+$ eine Lösung von (3.5). Ein weiteres Ausmultiplizieren des Terms ist nicht empfehlenswert, weil in dieser Form $(U\Sigma)^+$ nur einmal je Minimierung berechnet werden muss und lediglich die Pseudoinverse einer $s \times s$-Matrix in jedem Zielfunktionsaufruf zu bestimmen ist. Zusammengefasst wird die folgende Zielfunktion definiert:

$$F_{\text{hard}}(k) := F_{\text{soft}}\left(((U\Sigma)^+ C^{\text{dgl}}(k))^+, k\right). \tag{3.6}$$

Für den weiteren Verlauf dieser Arbeit wird für die Bestimmung einer durch ein kinetisches Modell regularisierten Matrixfaktorisierung die Minimierung von $F_{\text{hard}}(k)$ genutzt.

Zur Lösung des Anfangswertproblems werden in Matlab die Funktionen *ode45* oder *ode15s*, basierend auf [32, 37, 114], verwendet. Alternativ existieren für große Teile der Programme dieser Arbeit auch C Implementierungen, welche die C/Fortran Prozedur *radau5* nach [53, 130] verwenden. Die Optimierung wird in Matlab durch die Funktion *lsqnonlin* [22] beziehungsweise als C/Fortran Implementierungen mit *nl2sol*[5] [29] realisiert.

[5] https://dl.acm.org/citation.cfm?id=355966

3.2. Daten- und Modell-defizitäre Probleme

Im Idealfall entsprechen die Komponenten, welche durch ein kinetisches Modell beschrieben werden, exakt denen, die durch geeignete Linearkombinationen der Singulärvektoren von D rekonstruiert werden können. Es gilt also $\mathrm{Im}(U) = \mathrm{Im}(C^{\mathrm{dgl}}(k))$. Repräsentiert das Modell darüber hinaus weitere Komponenten, handelt es sich um ein *Daten-defizitäres* und bei einer geringeren Anzahl an Komponenten um ein *Modell-defizitäres* Problem.

Es stellt sich nun die Frage, ob eine sinnvolle Anpassung einer Kinetik auch noch für diese beiden Spezialfälle möglich ist. Sei hierzu s weiterhin durch $s = \mathrm{rg}(D)$ definiert und z die Anzahl der durch die Kinetik beschriebenen Komponenten.

Modell-defizitäre Probleme mit $s > z$ werden als erstes betrachtet. Weil $(U\Sigma) \in \mathbb{R}^{m\times s}$ und $C^{\mathrm{dgl}}(k) \in \mathbb{R}^{m\times z}$ dann nicht die gleichen Dimensionen haben, kann die Differenz zwischen Modell und Faktor nicht ohne Weiteres bestimmt werden, vergleiche (3.5). Es ist eine geeignete Ergänzung der Spalten der Matrix $T^p(k) := (U\Sigma)^+ C^{\mathrm{dgl}}(k)$ vorzunehmen. Die Transformationsmatrix lässt sich in Abhängigkeit der ergänzten Spalten $\bar{T}^p \in \mathbb{R}^{s\times(s-z)}$ und k durch

$$T(\bar{T}^p, k) := (T^p(k) \quad \bar{T}^p)^+$$

darstellen. Die Zielfunktion (3.4) wird wie folgt angepasst:

$$F_{\mathrm{hard},s>z}(\bar{T}^p, k) := F(T(\bar{T}^p, k)) + \delta_1 \sum_{i=1}^{m}\sum_{j=1}^{z} \left(\frac{(U\Sigma T^p(k))_{i,j} - (C^{\mathrm{dgl}}(k))_{i,j}}{\max_l((U\Sigma T^p(k))_{l,j})} \right)^2 .$$

Für die nicht durch das kinetische Modell beschriebenen Spalten des Faktors C, nämlich $U\Sigma\bar{T}^p$, ist zur Vermeidung trivialer Lösungsmengen eine Skalierung einzuführen. Es kann analog zu Abschnitt 2.3 auf die Perron-Frobenius Theorie zurückgegriffen und $(\bar{T}^p)_{1,j} = 1$ für $j = 1, \ldots, s - z$ festgelegt werden.

Als nächstes wird der Fall eines Daten-defizitären Problems mit $s < z$ untersucht. Hierzu sei $I \subset \{1, \ldots, z\}$ die Menge an Indizes derjenigen Spalten von $C^{\mathrm{dgl}}(k)$, die sich auch durch Linearkombinationen der linksseitigen Singulärvektoren von D rekonstruieren lassen. Durch Streichen der übrigen Spalten lässt sich wie für den Fall $s = z$ eine Transformationsmatrix bestimmen. Eine Anpassung der Zielfunktion (3.4) ist mit $T_I^p(k) := (U\Sigma)^+ (C^{\mathrm{dgl}}(k))_{:,I}$ in folgender Form möglich:

$$F_{\mathrm{hard},s<z}(k) := F((T_I^p(k))^+) + \delta_1 \sum_{i=1}^{m}\sum_{j\in I} \left(\frac{(U\Sigma T_I^p(k))_{i,j} - (C^{\mathrm{dgl}}(k))_{i,j}}{\max_l((U\Sigma T_I^p(k))_{l,j})} \right)^2 .$$

Auch eine Kombination von $F_{\mathrm{hard},s>z}(\bar{T}_z^p, k)$ und $F_{\mathrm{hard},s<z}(k)$ kann genutzt werden, wenn weder $\mathrm{Im}(U) \subseteq \mathrm{Im}(C^{\mathrm{dgl}})$ noch $\mathrm{Im}(C^{\mathrm{dgl}}) \subseteq \mathrm{Im}(U)$ gelten. Ein Beispiel für die Nutzung der Zielfunktion $F_{\mathrm{hard},s<z}(k)$ ist in Anhang A.2 zu finden.

3.3. Nichtnegativitätseigenschaft kinetischer Modelle

In diesem Abschnitt wird bewiesen, dass die Komponenten der vektorwertigen Lösung $c(t)$ eines kinetischen Modells (3.3) nur dann null werden können, wenn $t = t_0 = 0$ ist.

Der Beweis erfolgt in zwei Schritten. Zunächst wird in Hilfssatz 1 die Aussage

$$c_i(t) > 0 \quad \Rightarrow \quad c_i(t') > 0 \quad \text{für} \quad i \in \{1, \ldots, s\} \quad \text{und alle} \quad t' \geq t$$

gezeigt. Die Hauptaussage

$$c_i(t) > 0 \quad \text{für} \quad i \in \{1, \ldots, s\} \quad \text{und alle} \quad t > 0$$

wird anschließend in Satz 3 bewiesen.

In beiden Sätzen wird eine alternative Darstellung von $\dot{c}(t) = M(k)f(c(t))$ des Anfangswertproblems (3.3) genutzt. Durch die Polynome p_i, q_i mit $i = 1, \ldots, s$ erfolgt die additive Zerlegung der Ableitungen $\dot{c}_i(t)$ in positive und negative Summanden. Es wird vorausgesetzt, dass die negativen Summanden der i-ten Komponente $\dot{c}_i(t)$ stets den Faktor $c_i(t)$ enthalten. Dies folgt aus der speziellen Struktur kinetischer Modelle.

Hilfssatz 1. *Gegeben sei ein Anfangswertproblem*

$$\dot{c}(t) = \begin{pmatrix} p_1(c(t)) - q_1(c(t)) \cdot c_1(t) \\ \vdots \\ p_s(c(t)) - q_s(c(t)) \cdot c_s(t) \end{pmatrix}, \quad c(0) = c_0 \geq 0, \quad t \in [0, \infty) \tag{3.7}$$

mit $c : \mathbb{R}_{\geq 0} \to \mathbb{R}^s_{\geq 0}$ *und den Polynomen* $p_i, q_i : \mathbb{R}^s_{\geq 0} \to \mathbb{R}_{\geq 0}$ *der Form* $p_i(c(t)) := p_i(c_1(t), \ldots, c_s(t))$, $i = 1, \ldots, s$ *(q_i analog). Weiter sei $c(t)$ eine Lösung von* (3.7)*. Gilt für einen Index l und ein $t^+ \in \mathbb{R}_{\geq 0}$ die Ungleichung $c_l(t^+) > 0$, dann ist $c_l(t) > 0$ für alle $t \in [t^+, \infty)$.*

Beweis: Die Beweisführung erfolgt indirekt. Hierzu wird die Annahme, dass ein $t' > t^+$ mit $c_l(t') = 0$ existiert, zum Widerspruch geführt. Das Anfangswertproblem

$$\dot{d}(t) = p_l(c(t)) - q_l(c(t)) \cdot d(t), \quad d(t^+) = c_l(t^+), \quad \forall t \in [t^+, \infty)\,,$$

wird offensichtlich durch die Funktion $d(t) = c_l(t)$ gelöst. Die Lösung kann in den homogenen und inhomogenen Anteil

$$d(t) = d_H(t) + d_I(t) = \underbrace{d(t^+) \exp\left(\int_{t^+}^{t} -q_l(c(s))\,\mathrm{d}s\right)}_{>0} + \int_{t^+}^{t} p_l(c(s))\,\mathrm{d}s$$

zerlegt werden. Nach Voraussetzung ist $p_l(c(s)) \geq 0$ und damit auch $d_I(t) \geq 0$ für $t \in [t^+, \infty)$. Es folgt für $t \in [t^+, \infty)$, dass die Ungleichung $d(t) \geq d_H(t) > 0$ gilt und insbesondere ergibt sich mit $t' > t^+$ der Widerspruch $0 < d(t') = c_l(t') = 0$.

□

Um zwischen der kontinuierlichen Lösung $c(t)$ und ihrer Auswertung auf einem diskreten Zeitgitter zu unterscheiden, wird nun der folgende Operator eingeführt. Der Beweis der Hauptaussage dieses Abschnitts erfolgt anschließend.

Definition. Sei ein Gitter $T_G = (t_1, \ldots, t_m)^T$ bekannt. Der Operator $\mathcal{T}$ bildet die Funktion $c : \mathbb{R} \to \mathbb{R}^s$ auf den Vektor der Funktionsauswertungen von c zu den Stützstellen in T_G ab und ist wie folgt definiert:

$$\mathcal{T}c(t) := (c(t_1), \ldots, c(t_m))^T \in \mathbb{R}^{m \times s}\,.$$

Satz 3. *Gegeben sei das Anfangswertproblem* (3.7) *mit analog zu Hilfssatz 1 definierten Polynomen p_i, q_i für $i = 1, \ldots, s$ und der Lösung c. Weiter sei* $\mathrm{rg}(\mathcal{T}c(t)) = s$*. Dann ist $c(t) > 0$ für alle $t > 0$.*

Beweis:

Es seien zwei Indexmengen

$$I_+ = \{i \in \{1, \ldots, s\} : c_i(t) > 0 \ \forall t > 0\} \quad \text{und} \quad I_0 = \{1, \ldots, s\}/I_+$$

definiert. Für alle $c_i(t)$ mit $i \in I_+$ gilt mit Hilfssatz 1 die Behauptung. Für alle $i \in I_0$ existiert ein $\tau_i > 0$, sodass $c_i(t) = 0$ für $t \in [0, \tau_i]$ gilt. Sei weiter $\tau_{\min} := \min_{i \in I_0} \tau_i$ mit $\tau_{\min} > 0$. Unter diesen Vorbetrachtungen lautet die Behauptung des Satzes, dass $I_0 = \emptyset$ ist. Zum Beweis wird die komplementäre Aussage $I_0 \neq \emptyset$ zum Widerspruch geführt.

Seien also $l \in I_0$ und die l-te Komponente von $\dot{c}(t)$ aus (3.7) gegeben durch

$$\dot{c}_l(t) = p_l(c(t)) - q_l(c(t)) \cdot c_l(t)\,. \tag{3.8}$$

Es wird nun das Polynom $p_l(c(t))$ betrachtet und additiv in $p_l^+(c(t))$ und $p_l^0(c(t))$ zerlegt. Sei hierzu ein Summand von $p_l(c(t))$ durch $r(c(t)) = \alpha \prod_{j \in J} c_j(t)$ mit $\alpha > 0$ gegeben. Ist $J \cap I_0 = \emptyset$ wird r dem Polynom $p_l^+(c(t))$ und andernfalls $p_l^0(c(t))$ zugeordnet. Das Polynom $p_l^+(c(t))$ enthält also nur solche Summanden, die sich für $t > 0$ aus einem Produkt mit (echt) positiven Faktoren zusammensetzen. Für ein $t' > 0$ gilt also $p_l^+(c(t')) = 0$ genau dann, wenn p_l^+ das Nullpolynom ist.

Für $t \in [0, \tau_{\min}]$ und $i \in I_0$ gilt $c_i(t) = 0$. Folglich gelten für $l \in I_0$ auch $\dot{c}_l(t) = c_l(t) = 0$. Damit lässt sich (3.8) für $t \in [0, \tau_{\min}]$ durch

$$0 = c_l(t) = p_l^+(c(t)) \tag{3.9}$$

darstellen, weil die restlichen Summanden von (3.8) den Faktor 0 enthalten. Mit (3.9) gilt die Gleichung $0 = p_l^+(c(\tau_{\min}))$ für ein $t' = \tau_{\min} > 0$ und es folgt, dass p_l^+ das Nullpolynom ist. Damit kann (3.8) auf

$$\dot{c}_l(t) = p_l^0(c(t)) - q_l(c(t)) \cdot c_l(t) \tag{3.10}$$

reduziert werden.

Die Gleichung $p_l^0(c(t)) = 0$ gilt lediglich für $t \in [0, \tau_{\min}]$, aber nicht notwendigerweise für $t > \tau_{\min}$. Jeder Summand der rechten Seite von (3.10) enthält aber ein Element der Menge $\{c_i(t) : \ i \in I_0\}$ als Faktor. Zusammen mit den Anfangswerten $(c_0)_i = 0$ für $i \in I_0$ folgt $\dot{c}_l(t) = 0$ und somit, dass $c_l(t)$ konstant ist. Weil $l \in I_0$ ist, folgt insbesondere $c_l(t) = 0$ für alle $t > 0$. Dies widerspricht der Voraussetzung $\mathrm{rg}(\mathcal{T}c(t)) = s$ des Satzes und die Behauptung folgt.

□

Aus Satz 3 folgt die Hauptaussage dieses Abschnitts, dass $c_i(t) > 0$ für alle i und $t > 0$ gilt.

3.4. Numerische Beispiele

In diesem Abschnitt werden für zwei Datensätze regularisierte Matrixfaktorisierungen unter Zuhilfenahme kinetischer Modelle bestimmt. Der Fokus liegt hierbei auf der Bestimmung optimaler Parameter dieser Modelle. Als Überleitung für das folgende Kapitel 4 wird abschließend die Fragestellung nach dem Auftreten von nichttrivialen Lösungsmengen der regularisierten Matrixfaktorisierung trotz der Verwendung von Kinetiken anhand der beiden Beispiele untersucht.

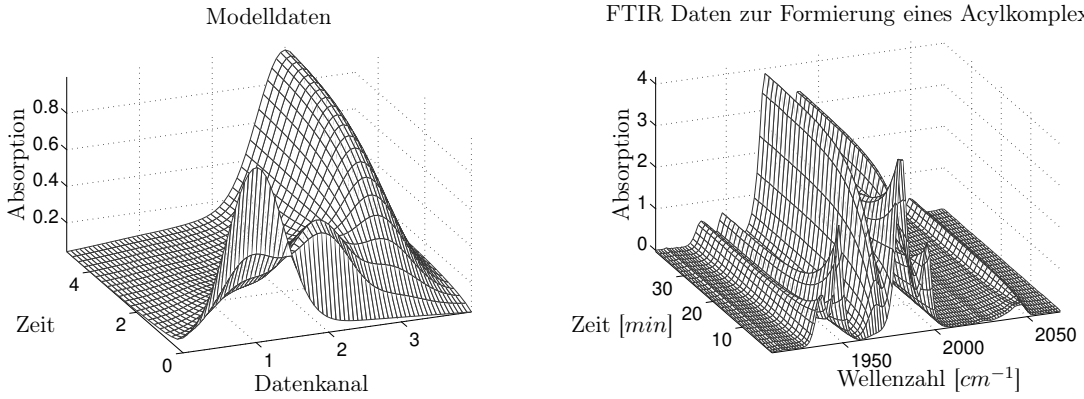


Abbildung 3.1.: Darstellung der Matrizen D des Modelldatensatzes 3 (links) und des spektroskopischen Datensatzes 1 (rechts).

Modelldatensatz 3. Es wird ein simulierter Datensatz mit $m = 101$ Spektren zu jeweils $n = 401$ Datenkanälen mit $s = 3$ Komponenten betrachtet. Der Vektor der Anfangskonzentrationen lautet $c_0 := (c_{\mathcal{X}}(0), c_{\mathcal{Y}}(0), c_{\mathcal{Z}}(0))^T = (1, 0, 0)^T$. Die zur Modellierung des Faktors C verwendete Kinetik ist

$$\mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \xrightarrow{k_3} \mathcal{Z} \tag{3.11}$$

mit Parametern $k_1 = 4$ und $k_2 = k_3 = 1$. Der Faktor S und die Diskretisierung der Zeitachse sind identisch mit Modelldatensatz 1. Die Matrix D ist in der linken Grafik der Abbildung 3.1 dargestellt.

Zur Bestimmung einer regularisierten Matrixfaktorisierung von D aus Modelldatensatz 3 wird das kinetische Modell (3.11) und die Zielfunktion (3.6) genutzt. Die Optimierung wird mit $k^{(0)} = (0.5, 0.5, 0.5)^T$ initialisiert und $\gamma_1 = \gamma_2 = \gamma_3 = \delta_1 = 1$ gewählt. Für die optimierten Parameter $k^* = (3.95, 1.04, 1.01)^T$ lautet der Zielfunktionswert $1.07 \cdot 10^{-18}$ und die ermittelten Faktoren sind in Abbildung 3.2 dargestellt. Trotz eines Zielfunktionswerts in der Größenordnung der Rechengenauigkeit ist auffällig, dass k^* nicht mit den ursprünglichen Geschwindigkeitsparametern $k_1 = 4$ und $k_2 = k_3 = 1$, welche zur Datengenerierung genutzt wurden, übereinstimmt. Ihre Rekonstruktion ist auf Basis der Minimierung von $F_{\text{hard}}(k)$ offensichtlich nicht möglich.

Spektroskopischer Datensatz 1 (Formierung eines Acylkomplexes)**.** Dieses Reaktionssystem ist in das Themengebiet der Hydroformylierung einzuordnen und beschreibt die Bildung eines Acylkomplexes aus zwei stereoisomeren Hydridokomplexen. Das entsprechende Katalysatorsystem ist in [69] erläutert. Die Versuchsdurchführung erfolgte durch Dr. Christoph Kubis am LIKAT[6] in Rostock. Es handelt sich um eine zeitaufgelöste Messfolge von $m = 735$ FTIR-Spektren zu jeweils $n = 325$ Wellenzahlen (Datenkanälen). Im zugrunde liegenden Frequenzfenster $[1911\text{cm}^{-1}, 2067\text{cm}^{-1}]$ absorbieren $s = 2$ Komponenten. Bei dem Reaktanten handelt es sich um eine Mischung aus zwei stereoisomeren Hydridokomplexen, die sich in einem schnellen Gleichgewicht befinden und folglich zu einer Komponente $\mathcal{E}$ zusammengefasst werden. Das Produkt $\mathcal{P}$ beschreibt den *Acylkomplex*. Es werden relative Konzentrationen betrachtet, womit der Vektor der Anfangskonzentrationen als $c_0 := (c_{\mathcal{E}}(0), c_{\mathcal{P}}(0))^T = (1, 0)^T$ angenommen werden kann. Die

[6] Leibniz-Institut für Katalyse e. V.

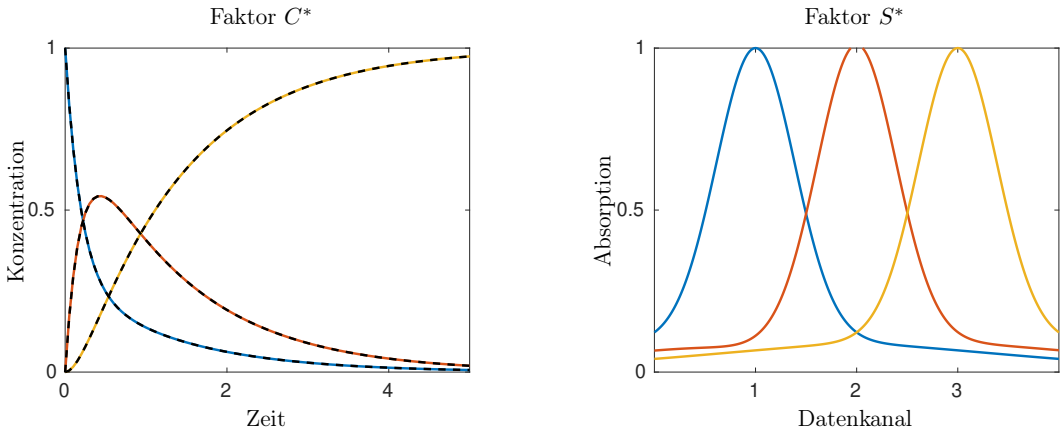


Abbildung 3.2.: Darstellung der Faktoren C^* und S^* zum optimierten Parameter k^* für den Modelldatensatz 3. Links sind die Spalten des Konzentrationsfaktors C^* als durchgängige farbige Linien sowie die Auswertung des kinetischen Modells zu k^* als gestrichelte schwarze Linien und rechts die Spalten des Spektrenfaktors S^* dargestellt.

Konzentrationen entsprechen dann dem prozentualen Anteil von $\mathcal{E}$ und $\mathcal{P}$ am Gesamtgemisch. Eine Darstellung der Matrix D ist in der rechten Grafik der Abbildung 3.1 zu sehen.

Die Konzentrationsverläufe der Komponenten $\mathcal{E}$ und $\mathcal{P}$ im spektroskopischen Datensatz 1 können mit dem Modell

$$\mathcal{E} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{P}$$

approximiert werden. Ausgehend von $k^{(0)} = (0.07\ \text{min}^{-1}, 0.07\ \text{min}^{-1})^T$ wird die Zielfunktion $F_{\text{hard}}(k)$ mit $\gamma_1 = \gamma_2 = \gamma_3 = \delta_1 = 1$ minimiert. Für die optimierten Parameter $k^* = (0.13\ \text{min}^{-1}, 0.047\ \text{min}^{-1})^T$ lautet der Zielfunktionswert $F_{\text{hard}}(k^*) = 0.0083$ und die weiteren Ergebnisse sind in Abbildung 3.3 dargestellt. Sowohl der Faktor C^* als auch S^* sind im zugrunde liegenden Kontext plausibel, wodurch von einer erfolgreichen Zerlegung ausgegangen werden kann.

Nichttriviale Lösungsmengen trotz kinetischer Modellierung

Für den Modelldatensatz 3 und den spektroskopischen Datensatz 1 sind im vorherigen Abschnitt Lösungen der jeweiligen regularisierten Matrixfaktorisierung dargestellt. Es wird nun die Existenz weiterer Lösungen untersucht.

Für den Modelldatensatz 3 wird eine Variation der Initialisierung der Geschwindigkeitsparameter $k^{(0)}$ vorgenommen. Die entsprechenden Werte können der Tabelle 3.1 entnommen werden. Analog zum vorherigen Abschnitt wird eine Minimierung der Zielfunktion $F_{\text{hard}}(k)$ durchgeführt. Die ebenfalls in der Tabelle zu findenden Werte der Zielfunktion für die jeweils optimalen Geschwindigkeitsparameter k^* sind nach Straffunktionen und Fehler der kinetischen Anpassung getrennt. Sowohl $F(T^*)$ als auch $F^{\text{dgl}}(C^*, k^*)$ aus Zielfunktion (3.4) liegen im Bereich der geforderten Genauigkeit der Optimierung und dennoch existieren starke Unterschiede in den optimierten Geschwindigkeitsparametern. Die grafische Darstellung der Parameter in Abbildung 3.4 lässt einen Zusammenhang zwischen den Komponenten der Vektoren k^* vermuten. In der mittleren und rechten

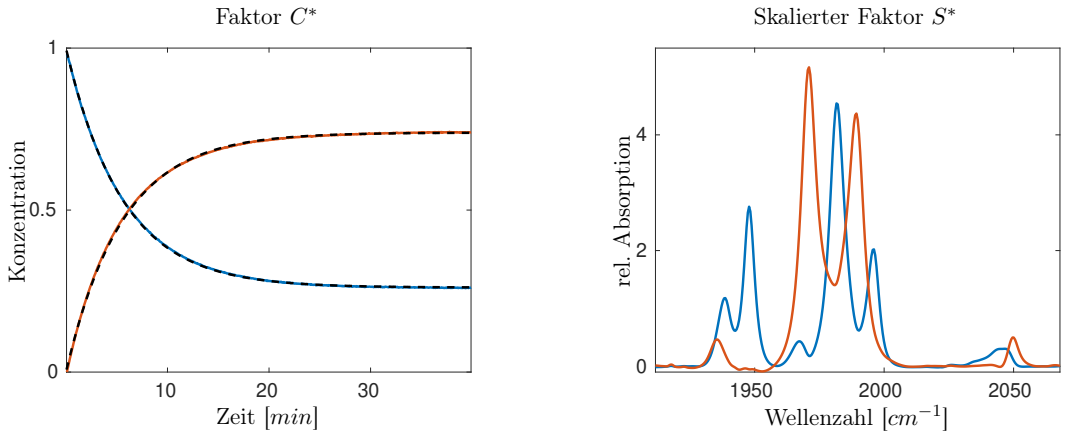


Abbildung 3.3.: Darstellung der Faktoren C^* und S^* der optimierten Parameter k^* zum spektroskopischen Datensatz 1. Links sind die Spalten des Konzentrationsfaktors C^* als durchgängige farbige Linien sowie die Auswertung des kinetischen Modells zu k^* als gestrichelte schwarze Linien und rechts die Spalten des Spektrenfaktors S^* dargestellt.

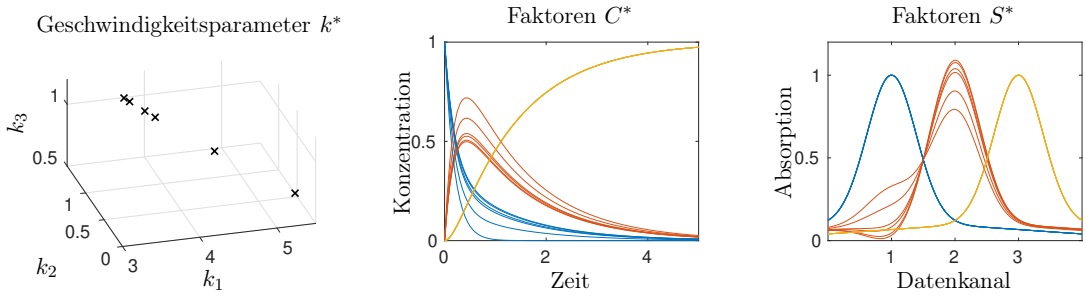


Abbildung 3.4.: Darstellung der Ergebnisse zum Modelldatensatz 3. Links sind die optimierten Geschwindigkeitsparameter zu verschiedenen Initialisierungen der Optimierung dargestellt. Die entsprechenden Faktoren C^* und S^* sind mittig und rechts zu sehen. Wegen des geringen Fehlers der kinetischen Anpassung wurde auf die Darstellung des kinetischen Modells in der mittleren Grafik verzichtet.

Grafik von Abbildung 3.4 sind sehr gut die qualitativen Unterschiede der zugehörigen Faktoren C^* und S^* zu sehen.

Eine ganz ähnliche Beobachtung kann für den spektroskopischen Datensatz 1 gemacht werden. Hierzu wird der Gewichtungsfaktor δ_1 der kinetischen Anpassung variiert. In Tabelle 3.2 ist zu sehen, dass mit der Variation von δ_1 auch geringe Änderungen von $F(T^*)$ und $F^{\text{dgl}}(C^*, k^*)$ einhergehen. Nichtsdestotrotz sind diese Werte in Bezug auf das Anwendungsproblem als klein einzustufen. Die Veranschaulichung der Faktoren C^* (mittig) und S^* (rechts) in Abbildung 3.5 unterstützt diese Aussage. In diesem Fall ist für die berechneten Vektoren k^* ein linearer Zusammenhang zu vermuten.

Für zwei regularisierte Matrixfaktorisierungsprobleme konnte gezeigt werden, dass eine eindeutige Faktorisierung trotz Regularisierung mit einem kinetischen Modell nicht immer möglich ist. In Kapitel 4 erfolgt eine detaillierte Analyse dieses Zusammenhangs für den idealisierten Fall ungestörter Matrizen D. Darauf aufbauend erfolgt in Kapitel 5 eine Störungsanalyse.

Modelldatensatz 3	Initialisierung			optimiert			$F(T^*)$	$F^{\mathrm{dgl}}(C^*, k^*)$
	$k_1^{(0)}$	$k_2^{(0)}$	$k_3^{(0)}$	k_1^*	k_2^*	k_3^*		
	1	1	1	3.93	1.055	1.019	$2.26 \cdot 10^{-10}$	$1.03 \cdot 10^{-9}$
	1	1	3	4.49	0.62	0.89	$2.65 \cdot 10^{-10}$	$1.14 \cdot 10^{-9}$
	1	3	1	3.68	1.23	1.09	$2.09 \cdot 10^{-10}$	$9.92 \cdot 10^{-10}$
	3	1	1	3.83	1.13	1.05	$2.20 \cdot 10^{-10}$	$1.01 \cdot 10^{-9}$
	3	3	3	3.63	1.27	1.1	$2.05 \cdot 10^{-10}$	$9.81 \cdot 10^{-10}$
	5	0	1	5.23	0.001	0.76	$3.22 \cdot 10^{-10}$	$1.30 \cdot 10^{-9}$

Tabelle 3.1.: Übersicht der Ergebnisse zum Modelldatensatz 3 für verschiedene Initialisierungen der Geschwindigkeitsparameter $k^{(0)}$.

Spek. Datensatz 1		Initialisierung		optimiert			
	δ_1	$k_1^{(0)}$	$k_2^{(0)}$	k_1^*	k_2^*	$F(T^*)$	$F^{\mathrm{dgl}}(C^*, k^*)$
	0.05	0.1	0.1	0.154	0.026	0.17	0.10
	0.1	0.1	0.1	0.151	0.029	0.18	0.099
	0.2	0.1	0.1	0.146	0.034	0.20	0.096
	0.4	0.1	0.1	0.139	0.04	0.24	0.093
	0.8	0.1	0.1	0.134	0.046	0.33	0.09

Tabelle 3.2.: Übersicht der Ergebnisse zum spektroskopischen Datensatz 1. Die Summanden der Zielfunktion sind separiert nach Straffunktionen und kinetischer Anpassung. Sie wurden in Abhängigkeit von δ_1 bestimmt.

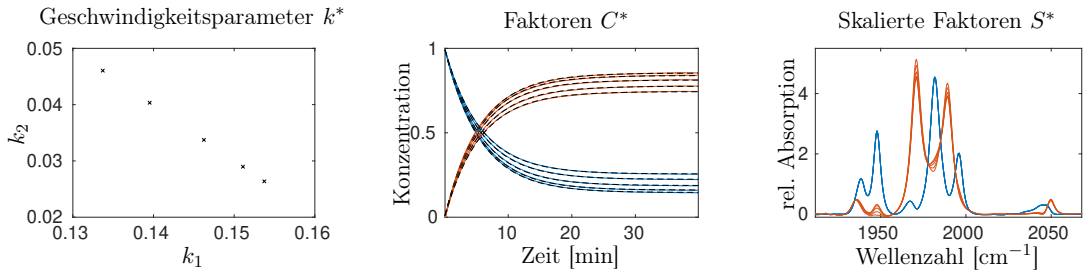


Abbildung 3.5.: Darstellung der Ergebnisse zum spektroskopischen Datensatz 1. Links sind die optimierten Geschwindigkeitsparameter zu verschiedenen δ_1 dargestellt. Die entsprechenden Faktoren C^* und S^* sind mittig und rechts zu sehen. Für den Faktor C^* sind zusätzlich die jeweiligen Auswertungen des kinetischen Modells als schwarz gestrichelte Linie dargestellt.

4. Lösungsmengen von Kinetikparametrierungen

In diesem Kapitel wird die Existenz von Mengen kinetischer Parametrierungen untersucht, die bezüglich einer gegebenen Matrix D eine nichtnegative Vollrangfaktorisierung $D = CS^T$ bei gleichzeitiger Konsistenz des Faktors C mit einer Kinetik ermöglichen. In der Literatur [44, 125, 128] ist das allgemeine Problem der Eindeutigkeit von Modellparametern unter dem englischen Stichwort "Identifiability" bekannt, also der Identifizierbarkeit von Parametern. Die ersten analytischen Betrachtungen zu einer einfachen Folgereaktion wurden 1970 durchgeführt [1]. Die verwendeten Datensätze basieren auf zeitabhängigen UV/Vis-spektroskopischen Messungen zu nur einer Wellenlänge und der Fokus liegt nicht auf der Lösung eines Matrixfaktorisierungsproblems, sondern auf der Anpassung eines kinetischen Modells. Die Autoren stellten fest, dass durch Permutation der optimierten Geschwindigkeitsparameter qualitativ gleichwertige Anpassungen des kinetischen Modells erzielt werden können. Die Einführung des hierfür bis heute gebräuchlichen Begriffs der "slow-fast ambiguity" erfolgt in [62]. Die Arbeiten [19, 84, 118] enthalten weiterführende Untersuchungen zu ausgewählten komplexeren Kinetiken. Eine erste allgemeine Beschreibung der Lösungskontinua kinetischer Parameter für beliebige Kinetiken im Kontext der Spektrenanalyse geht auf Vajda und Rabitz zurück [125, 126]. Dies sind zentrale Arbeiten des Forschungsgebiets. Sie greifen erstmals die Idee der linearen Transformation der Matrix $C^{\mathrm{dgl}}(k)$, siehe Abschnitt 3.1, zur Beschreibung der Parameterlösungsmengen auf. Eine solche Transformation kann als Analogon der Matrizen T in (2.2) zur Beschreibung der Menge zulässiger Lösungen verstanden werden. In [116] wird erstmals die "slow-fast ambiguity" von zweistufigen Folgereaktionen im Kontext multivariater spektroskopischer Methoden untersucht. Eine Ausweitung auf weitere kinetische Modelle erfolgt in [3, 64]. Die Definition einer Lösungsmenge kinetischer Parameter für beliebige kinetische Modelle erster Ordnung durch die Eigenwerte der Koeffizientenmatrix des zugeordneten gewöhnlichen Differentialgleichungssystems ist in [112] zu finden. Hierdurch wird erstmals eine effiziente numerische Berechnung dieser Lösungskontinua ermöglicht.

Dieses Kapitel baut insbesondere auf den Ergebnissen in [112] auf. In Abschnitt 4.1 werden Parameterlösungsmengen von kinetischen Modellen erster Ordnung detailliert analysiert. In einem beschränkten Maße treten nichttriviale Lösungsmengen von Kinetikparametrierungen auch für Kinetiken zweiter Ordnung auf, was in Abschnitt 4.2 untersucht wird.

Eine Kinetik beliebiger Ordnung kann analog zu (3.3) mittels eines Anfangswertproblems

$$\dot{c}(t) = m(c(t), k), \quad c(t_0) = c_0 \geq 0 \tag{4.1}$$

dargestellt werden. In Beispiel 2 sind die durch $m(c(t), k)$ definierten Differentialgleichungssysteme für zwei ausgewählte Kinetiken gezeigt.

Beispiel 2. *Für das kinetische Modell erster Ordnung*

$$\mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y}$$

genügen die Konzentrationsverläufe dem gewöhnlichen Differentialgleichungssystem

$$\dot{c}(t) = \begin{pmatrix} \dot{c}_{\mathcal{X}}(t) \\ \dot{c}_{\mathcal{Y}}(t) \end{pmatrix} = \begin{pmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix} \begin{pmatrix} c_{\mathcal{X}}(t) \\ c_{\mathcal{Y}}(t) \end{pmatrix} =: m(c(t), k)\,.$$

Analoges gilt für das kinetische Modell zweiter Ordnung

$$\mathcal{X} + \mathcal{Y} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Z}$$

mit

$$\dot{c}(t) = \begin{pmatrix} \dot{c}_{\mathcal{X}}(t) \\ \dot{c}_{\mathcal{Y}}(t) \\ \dot{c}_{\mathcal{Z}}(t) \end{pmatrix} = \begin{pmatrix} -k_1 & k_2 \\ -k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix} \begin{pmatrix} c_{\mathcal{X}}(t) \cdot c_{\mathcal{Y}}(t) \\ c_{\mathcal{Z}}(t) \end{pmatrix} =: m(c(t), k)\,.$$

In Abschnitt 3.4 wurde bereits exemplarisch gezeigt, dass das nichtnegative Matrixfaktorisierungsproblem selbst mit der Regularisierung durch ein kinetisches Modell nicht eindeutig lösbar sein muss. Die Idee der niedrigdimensionalen Darstellung von Faktorisierungen durch die Menge zulässiger Lösungen in Abschnitt 2.3 wird erneut angewendet. Eine nichtnegative Faktorisierung, die ebenfalls konsistent mit einer Kinetik ist, ist bereits durch die Parameter $k \in \mathbb{R}^q$ der Kinetik eindeutig bestimmt. Analog zu den Vorüberlegungen zu Zielfunktion (3.6) können C und S mit $T = ((U\Sigma)^+ C^{\mathrm{dgl}}(k))^+$ in Abhängigkeit von k bestimmt werden. Diejenigen Faktorisierungen, die konsistent mit der Kinetik sind, lassen sich dann im q-dimensionalen Raum darstellen, der durch die q Parameter der jeweiligen Kinetik aufgespannt wird. Analog zur Menge zulässiger Lösungen wird nun ein Zugang zur Bestimmung aller "relevanten" Kinetikparametrierungen erläutert. Solche Lösungsmengen von Parametern einer Kinetik sind das zentrale Objekt dieser Arbeit und werden im Folgenden definiert.

Definition 1 (Menge D-konsistenter Parameter [112])**.** Seien die Matrix D mit Rang $s = \mathrm{rg}(D) = \mathrm{rg}_+(D)$ und die abgeschnittene Singulärwertzerlegung $D = U\Sigma V^T$ gegeben. Die Menge der D-konsistenten Parameter ist definiert als

$$\begin{aligned} \mathcal{K} :=& \{k \in \mathbb{R}^q_+ : c(t) \text{ löst (4.1) mit } C = \mathcal{T}c(t) \wedge \\ & \exists T \in \mathbb{R}^{s\times s} : (\mathrm{rg}(T) = s \ \wedge \ C = U\Sigma T^{-1})\}. \end{aligned}$$

Definition (Menge zulässiger Parameter [112])**.** Seien die Voraussetzungen wie in Definition 1. Die Menge der zulässigen Parameter ist definiert als

$$\mathcal{K}^+ := \{k \in \mathcal{K} : S^T = TV^T \geq 0\}.$$

Die Menge D-konsistenter Parameter $\mathcal{K}$ beinhaltet also diejenigen Parameter k eines kinetischen Modells zu einer Matrix D, sodass eine (nicht notwendigerweise nichtnegative) Faktorisierung $D = CS^T$ existiert und gleichzeitig $C = C^{\mathrm{dgl}}(k)$ gilt. Die Menge zulässiger Parameter $\mathcal{K}^+$ ist diejenige Teilmenge von $\mathcal{K}$, für die auch die entsprechenden Faktoren S nichtnegativ sind und somit durch $D = CS^T$ eine nichtnegative Vollrangfaktorisierung

gegeben ist. In den folgenden Kapiteln beziehen sich diese Definitionen immer auf die aktuell zugrunde liegende Kinetik sowie das zugehörige Zeitgitter T_G.

Soll mittels Definition 1 überprüft werden, ob ein k zur Menge $\mathcal{K}$ gehört, entfällt ein Großteil des Aufwands auf die Bestimmung der Lösung des Anfangswertproblems (4.1), weil dieses häufig nur numerisch zu berechnen ist. Im folgenden Abschnitt 4.1 wird bewiesen, dass für Kinetiken erster Ordnung stattdessen das Betrachten eines $s \times s$ Eigenwertproblems ausreicht. Für die in dieser Arbeit relevanten Reaktionssysteme gilt $s \leq 5$, wodurch eine signifikante Reduktion des Aufwands erreicht wird.

4.1. Kinetiken erster Ordnung

Das Differentialgleichungssystem zur Beschreibung eines allgemeinen kinetischen Modells wie in Gleichung (4.1) lässt sich für kinetische Modelle erster Ordnung in eine Matrix-Vektor-Darstellung mit einer quadratischen parameterabhängigen Koeffizientenmatrix $M(k) \in \mathbb{R}^{s\times s}$ überführen:

$$\dot{c}(t) = m(c(t), k) = M(k)c(t), \quad c(t_0) = c_0. \tag{4.2}$$

Die zentralen Sätze 4 und 5 des Kapitels ermöglichen die Beschreibung der Menge $\mathcal{K}$ mittels der Eigenwerte der Matrix $M(k)$ unter der Voraussetzung, dass ein $k \in \mathcal{K}$ bekannt ist. In Kapitel 3 wurde gezeigt, wie ein solches initiales k berechnet werden kann. Der Satz 4 behandelt den einfacheren Fall diagonalisierbarer Matrizen $M(k)$. Der Fall nichtdiagonalisierbarer Matrizen $M(k)$ kann als Verallgemeinerung angesehen werden und wird in Satz 5 bewiesen.

Es wird mit einem Hilfssatz begonnen, der für allgemeine Matrizen $M(k)$ gültig ist und eine Aussage über die Struktur der Linearkombination von Eigen- und Hauptvektoren von $M(k)$ zur Darstellung der Initialkonzentrationen c_0 trifft.

Hilfssatz 2. *Seien die Matrix $M := M(k) \in \mathbb{R}^{s\times s}$ zum Anfangswertproblem (4.2) mit Lösung $c(t)$ sowie deren Eigenwerte $\lambda_1, \ldots, \lambda_p$ mit Vielfachheiten $\mu_1, \ldots, \mu_p$ gegeben. Seien $J = \operatorname{diag}(J_1, \ldots, J_p)$ die Jordanmatrix von M mit Jordanblöcken J_i zu λ_i für $i = 1, \ldots, p$ und X die Matrix der zugehörigen, spaltenweise angeordneten Eigen- und Hauptvektoren, sodass $M = XJX^{-1}$ gilt. Weiter sei $\alpha = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)^T \in \mathbb{R}^s$ mit $\tilde{\alpha}_i \in \mathbb{R}^{\mu_i}$ für $i = 1, \ldots, p$, sodass $c_0 = X\alpha$ gilt.*
Gilt $\operatorname{rg}(\mathcal{T}(c(t))) = s$, dann ist $(\tilde{\alpha}_i)_{\mu_i} \neq 0$ für $i = 1, \ldots, p$.

Beweis: Die Lösung eines Anfangswertproblems der Form (4.2) lässt sich durch

$$c(t) = \mathrm{e}^{Mt}c_0$$

darstellen. Es folgt

$$\begin{aligned} C = \mathcal{T}c(t) = \begin{pmatrix} c(t_1)^T \\ \vdots \\ c(t_m)^T \end{pmatrix} = \begin{pmatrix} (\mathrm{e}^{Mt_1}c_0)^T \\ \vdots \\ (\mathrm{e}^{Mt_m}c_0)^T \end{pmatrix} = \begin{pmatrix} (\mathrm{e}^{XJX^{-1}t_1}c_0)^T \\ \vdots \\ (\mathrm{e}^{XJX^{-1}t_m}c_0)^T \end{pmatrix} \\ = \begin{pmatrix} (\mathrm{e}^{XJX^{-1}t_1}X\alpha)^T \\ \vdots \\ (\mathrm{e}^{XJX^{-1}t_m}X\alpha)^T \end{pmatrix} = \begin{pmatrix} (\mathrm{e}^{Jt_1}\alpha)^T \\ \vdots \\ (\mathrm{e}^{Jt_m}\alpha)^T \end{pmatrix} X^T. \end{aligned}$$

Es sei $J_i = D_i + N_i$ die additive Zerlegung in den Haupt- und Nebendiagonalanteil des i-ten Jordanblocks. Es folgt für $i = 1, \ldots, p$, dass

$$\begin{aligned} \mathrm{e}^{J_i t}\tilde{\alpha}_i &= \mathrm{e}^{(D_i+N_i)t}\tilde{\alpha}_i = \mathrm{e}^{D_i t}\mathrm{e}^{N_i t}\tilde{\alpha}_i \\ &= \mathrm{e}^{\lambda_i t}\begin{pmatrix} 1 & \frac{t}{1!} & \cdots & \frac{t^{\mu_i-1}}{(\mu_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{t}{1!} \\ & & & 1 \end{pmatrix}\tilde{\alpha}_i = \mathrm{e}^{\lambda_i t}\underbrace{\begin{pmatrix} (\tilde{\alpha}_i)_1 + \frac{t}{1!}(\tilde{\alpha}_i)_2 + \ldots + \frac{t^{\mu_i-1}}{(\mu_i-1)!}(\tilde{\alpha}_i)_{\mu_i} \\ \vdots \\ (\tilde{\alpha}_i)_{\mu_i-1} + \frac{t}{1!}(\tilde{\alpha}_i)_{\mu_i} \\ (\tilde{\alpha}_i)_{\mu_i} \end{pmatrix}}_{v_i(t)} \end{aligned}$$

gilt. Mit der Blockgestalt von J und der vorherigen Gleichung folgt

$$\begin{aligned} C(X^T)^{-1} &= \begin{pmatrix} (\mathrm{e}^{Jt_1}\alpha)^T \\ \vdots \\ (\mathrm{e}^{Jt_m}\alpha)^T \end{pmatrix} = \begin{pmatrix} (\mathrm{e}^{J_1 t_1}\tilde{\alpha}_1)^T & \cdots & (\mathrm{e}^{J_p t_1}\tilde{\alpha}_p)^T \\ \vdots & & \vdots \\ (\mathrm{e}^{J_1 t_m}\tilde{\alpha}_1)^T & \cdots & (\mathrm{e}^{J_p t_m}\tilde{\alpha}_p)^T \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{e}^{\lambda_1 t_1}(v_1(t_1))^T & \ldots & \mathrm{e}^{\lambda_p t_1}(v_p(t_1))^T \\ \vdots & & \vdots \\ \mathrm{e}^{\lambda_1 t_m}(v_1(t_m))^T & \ldots & \mathrm{e}^{\lambda_p t_m}(v_p(t_m))^T \end{pmatrix}. \end{aligned} \tag{4.3}$$

Es wird nun angenommen, dass ein $i^* \in \{1, \ldots, p\}$ existiert, sodass $(\tilde{\alpha}_{i^*})_{\mu_{i^*}} = 0$ gilt. Damit folgt aber sofort, dass (4.3) nicht den vollen Spaltenrang hat, weil die letzte Komponente von $v_{i^*}(t)$ dann für alle t null ist. Dies steht im Widerspruch zum vollen Spaltenrang der Matrix $C(X^T)^{-1}$. Die Behauptung des Satzes folgt.

□

Bemerkung 1. Aus der Anwendung von Hilfssatz 2 auf eine diagonalisierbare Matrix $M(k)$ folgt, dass das entsprechende α keine Nulleinträge hat.

Es wird eine weitere Aussage über allgemeine reguläre Matrizen $A \in \mathbb{R}^{n\times n}$ benötigt, welche im folgende Hilfssatz bewiesen wird. Sie besagt, dass für A immer Vektoren x und y existieren, sodass x und y für Indizes einer gemeinsamen Indexmenge keine Nulleinträge aufweisen und $Ax = y$ gilt.

Hilfssatz 3. *Seien $A \in \mathbb{R}^{n\times n}$ regulär und $I \subseteq \{1, \ldots, n\}$.*
Dann existieren Vektoren $x, y \in \mathbb{R}^n$ mit $x_i \neq 0$ und $y_i \neq 0$ für $i \in I$, sodass $Ax = y$ gilt.

Beweis: Es seien $\mathbf{1}_n = (1, \ldots, 1)^T \in \mathbb{R}^n$ der Einsvektor und $e_i \in \mathbb{R}^n$ der i-te Einheitsvektor. Weiter seien $\tilde{y} := A\mathbf{1}_n$ und $J := \{i \in I : \tilde{y}_i = 0\}$ definiert. Die Menge J kann maximal $n-1$ Indizes enthalten, weil $\tilde{y}$ sonst der Nullvektor wäre und dies der Regularität von A widerspricht. Sei zudem $v^i := A^{-1}e_i$ für $i \in J$ definiert. Der durch

$$y = \tilde{y} - \underbrace{\frac{1}{(n-1)\max\limits_{i\in J}\|v^i\|_\infty}}_{\theta}\sum_{i\in J} e_i \tag{4.4}$$

definierte Vektor y enthält keine Nulleinträge mehr. Weiter folgen mit (4.4) die Gleichungen

$$y = A\mathbf{1}_n - \theta\sum_{i\in J} Av^i = A\underbrace{\left(\mathbf{1}_n - \theta\sum_{i\in J} v^i\right)}_{=:x}.$$

Der verwendete Vorfaktor θ skaliert die Komponenten der Summe $\sum_{i\in J} v^i$ so, dass diese betragsmäßig kleiner als 1 sind. Hierdurch gilt $x_i > 0$ für alle $i = 1, \ldots, n$ und somit insbesondere für $i \in I$, womit die Behauptung bewiesen ist.

□

Es wird weiter bewiesen, dass immer nur genau ein Hauptraum zu einem Eigenwert einer Matrix $M(k)$ wie in Hilfssatz 2 existieren kann.

Hilfssatz 4. *Sei $c(t)$ eine Lösung des Anfangswertproblems (4.2) mit $M := M(k)$. Weiter sei $\mathrm{rg}(\mathcal{T}c(t)) = s$ sowie (λ, v) ein Eigenpaar von M.*
Dann existiert kein Eigenpaar (λ, w) mit $v \nparallel w$.

Beweis: Es wird ein Widerspruchsbeweis durchgeführt. Die komplementäre Aussage sei, dass ein Eigenpaar (λ, w) mit $v \nparallel w$ existiert. Somit lässt sich die Lösung $c(t)$ als

$$c(t) = \alpha_1 \mathrm{e}^{\lambda t} v + \alpha_2 \mathrm{e}^{\lambda t} w + \sum_{i=3}^{s} \alpha_i y_i(t) = (\alpha_1 v + \alpha_2 w)\mathrm{e}^{\lambda t} + \sum_{i=3}^{s} \alpha_i y_i(t)$$

mit geeigneten $y_i : \mathbb{R} \to \mathbb{R}^n$ darstellen. Damit sind die Komponenten von $c_i(t)$ Linearkombinationen von maximal $s-1$ linear unabhängigen Funktionen. Hieraus folgt sofort, dass $\mathrm{rg}(\mathcal{T}c(t)) < s$ und somit der Widerspruch.

□

Es folgt nun der erste zentrale Satz zu den Eigenwerten der Koeffizientenmatrix $M(k)$ für $k \in \mathcal{K}$.

Satz 4. *Sei $D \in \mathbb{R}^{m\times n}$ eine nichtnegative Matrix mit $s = \mathrm{rg}(D) = \mathrm{rg}_+(D)$. Bezüglich der Matrix D sei der Vektor $k^* \in \mathbb{R}^q$ D-konsistent im Sinne der Definition 1 mit der zugehörigen Lösung des Anfangswertproblems $c^*(t)$. Seien weiter zu einem Zeitgitter $(t_1, \ldots, t_m)$ und einer regulären Matrix $T^* \in \mathbb{R}^{s\times s}$ die Gleichung $\mathcal{T}c^*(t) = C^* = U\Sigma(T^*)^{-1}$ erfüllt, sowie $k \in \mathbb{R}^q_+$ mit diagonalisierbaren Matrizen $M := M(k)$ und $M^* := M(k^*)$, siehe (4.2), gegeben.*
Dann ist k genau dann D-konsistent, wenn M und M^* ähnlich sind.

Beweis: "⇒": Mit $k, k^* \in \mathcal{K}$ gelten

$$\mathcal{T}c^*(t) = C^* = U\Sigma(T^*)^{-1} \text{ und } \mathcal{T}c(t) = C = U\Sigma T^{-1}\,.$$

Es folgt

$$\begin{aligned} CT = U\Sigma = C^*T^* \;&\Leftrightarrow C = C^* \underbrace{T^*T^{-1}}_{=:X} = C^*X \\ &\Leftrightarrow c(t)^T = c^*(t)^T X \quad \forall t \in \{t_1, \ldots, t_m\} \\ &\Leftrightarrow c(t) = X^T c^*(t) \quad \forall t \in \{t_1, \ldots, t_m\}\,. \end{aligned} \tag{4.5}$$

Als Lösung eines linearen Differentialgleichungssystems mit konstanter und diagonalisierbarer Koeffizientenmatrix lässt sich $c(t)$ als

$$c(t) = B \begin{pmatrix} \mathrm{e}^{\lambda_1 t} \\ \vdots \\ \mathrm{e}^{\lambda_s t} \end{pmatrix} \tag{4.6}$$

mit regulärer Matrix $B \in \mathbb{R}^{s\times s}$ und den, nach Hilfssatz 4, paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_s$ von M darstellen. Analog folgt die Darstellung für $c^*(t)$ mit regulärer Matrix $B^* \in \mathbb{R}^{s\times s}$ und den Eigenwerten $\lambda_1^*, \ldots, \lambda_s^*$ von M^*. Einsetzen in Gleichung (4.5) resultiert in

$$B \begin{pmatrix} \mathrm{e}^{\lambda_1 t} \\ \vdots \\ \mathrm{e}^{\lambda_s t} \end{pmatrix} = X^T B^* \begin{pmatrix} \mathrm{e}^{\lambda_1^* t} \\ \vdots \\ \mathrm{e}^{\lambda_s^* t} \end{pmatrix} \quad \forall t \in \{t_1, \ldots, t_m\}$$

$$\Leftrightarrow \begin{pmatrix} \mathrm{e}^{\lambda_1 t} \\ \vdots \\ \mathrm{e}^{\lambda_s t} \end{pmatrix} = \underbrace{B^{-1} X^T B^*}_{=:G} \begin{pmatrix} \mathrm{e}^{\lambda_1^* t} \\ \vdots \\ \mathrm{e}^{\lambda_s^* t} \end{pmatrix} \quad \forall t \in \{t_1, \ldots, t_m\}. \tag{4.7}$$

Für die folgende Betrachtung ist es wichtig zu bemerken, dass die Exponentialfunktionen der Menge $\{\mathrm{e}^{\lambda_1 t}, \ldots, \mathrm{e}^{\lambda_q t}\}$ mit $\lambda_i \neq \lambda_j$ für $i \neq j$ linear unabhängig sind, was leicht mittels der Wronski-Determinante für $t = 0$ nachgewiesen werden kann. Für die erste Komponente der Gleichung (4.7)

$$\mathrm{e}^{\lambda_1 t} = G(1,:) \begin{pmatrix} \mathrm{e}^{\lambda_1^* t} \\ \vdots \\ \mathrm{e}^{\lambda_s^* t} \end{pmatrix} \quad \forall t \in \{t_1, \ldots, t_m\}$$

folgt also, dass ein j^* existiert, sodass $\lambda_1 = \lambda_{j^*}^*$ gilt. Weil die Eigenwerte $\lambda_i,\ i = 1, \ldots, s$, paarweise verschieden sind, existiert nun zu jedem λ_i genau ein λ_j^*, sodass $\lambda_i = \lambda_j^*$ gilt. Die beiden diagonalisierbaren Matrizen M und M^* besitzen also die gleichen Eigenwerte und sind somit ähnlich zueinander.

"$\Leftarrow$": Seien (λ_i, x^i) mit $i = 1, \ldots, s$ die Rechtseigenpaare von M^* mit $X := (x^1, \ldots, x^s)$ und weiter (λ_i, y^i) mit $i = 1, \ldots, s$ die Linkseigenpaare von M sowie $Y := (y^1, \ldots, y^s)^T$. Damit gilt

$$X^{-1} M^* X = \mathrm{diag}(\lambda_1, \ldots, \lambda_s) =: D_M, \tag{4.8}$$

$$Y M Y^{-1} = D_M. \tag{4.9}$$

Es werden nun zwei zunächst beliebige reguläre Diagonalmatrizen $D_1, D_2 \in \mathbb{R}^{s\times s}$ eingeführt. Wird bezüglich der D-Konsistenz von k die Forderung nach Einhaltung der Anfangswerte $c(t_0) = c_0$ vernachlässigt, können D_1 und D_2 als Freiheitsgrade angesehen werden. Im abschließenden Teil des Beweises wird gezeigt, dass diese Freiheitsgrade so gewählt werden können, dass die Bedingung $c(t_0) = c_0$ erfüllt ist. Mit (4.8) und (4.9) folgt

$$\begin{aligned} M^* &= X D_M X^{-1} = X D_1 D_1^{-1} D_M X^{-1} = X D_1 D_M D_1^{-1} X^{-1} \\ &= \underbrace{X D_1}_{\tilde{X}} D_M (X D_1)^{-1} = \tilde{X} D_M \tilde{X}^{-1}, \\ M &= Y^{-1} D_M Y = Y^{-1} D_2^{-1} D_2 D_M Y = Y^{-1} D_2^{-1} D_M D_2 Y \\ &= (Y D_2)^{-1} D_M \underbrace{Y D_2}_{\tilde{Y}} = \tilde{Y}^{-1} D_M \tilde{Y}. \end{aligned}$$

Damit kann die Ähnlichkeitstransformation $M^* = \tilde{X}\tilde{Y}M(\tilde{X}\tilde{Y})^{-1} = ZMZ^{-1}$ mit $Z = \tilde{X}\tilde{Y}$ definiert werden und es gilt für das Anfangswertproblem

$$\begin{aligned} & \dot{c^*}(t) = M^*c^*(t), \quad c^*(t_0) = c_0 \\ \Leftrightarrow \quad & \dot{c^*}(t) = ZMZ^{-1}c^*(t), \quad c^*(t_0) = c_0 \\ \Leftrightarrow \quad & (\dot{Z^{-1}c^*})(t) = MZ^{-1}c^*(t), \quad c^*(t_0) = c_0\,. \end{aligned}$$

Es bleibt zu zeigen, dass D_1 und D_2 so gewählt werden können, dass $Z^{-1}c_0 = c_0$ gilt, womit die Funktion $c(t) := Z^{-1}c^*(t)$ schlussendlich das Anfangswertproblem (4.2) löst. Hierzu wird die Darstellung $c_0 = \tilde{X}\alpha$ von c_0 im Raum, der durch die Eigenvektoren $x^1, \ldots, x^s$ aufgespannt wird, genutzt. Es gelten die Äquivalenzaussagen

$$\begin{aligned} Z^{-1}c_0 = c_0 \quad & \Leftrightarrow \quad c_0 = Zc_0 \quad \Leftrightarrow \quad c_0 = \tilde{X}\tilde{Y}c_0 \quad \Leftrightarrow \quad \tilde{X}\alpha = \tilde{X}\tilde{Y}\tilde{X}\alpha \\ & \Leftrightarrow \quad \alpha = \tilde{Y}\tilde{X}\alpha \quad \Leftrightarrow \quad \alpha = D_2YXD_1\alpha \quad \Leftrightarrow \quad D_2^{-1}\alpha = YXD_1\alpha \\ & \Leftrightarrow \quad \beta = YX\gamma \end{aligned} \tag{4.10}$$

mit $\beta := D_2^{-1}\alpha$ und $\gamma := D_1\alpha$. Die Matrizen D_1 und D_2 sind bis auf die Forderung nach Regularität noch beliebig und nach Bemerkung 1 hat α keine Nullkomponenten. Es folgt, dass die Vektoren β und γ ebenfalls beliebig, aber komponentenweise ungleich Null sind. Es bleibt zu zeigen, dass solche Vektoren β und γ existieren, sodass die Gleichung (4.10) erfüllt ist. Dies folgt direkt aus Hilfssatz 3 mit $A = YX$. Es bleibt noch zu zeigen, dass $\mathrm{rg}(T) = s$ ist. Die Transformation $Zc(t) := c^*(t)$ gilt trivialerweise auch auf dem zugrunde liegenden Zeitgitter und es gelten

$$\begin{aligned} & U\Sigma(T^*)^{-1} = C^* = \mathcal{T}(c^*) = \mathcal{T}(Zc) = \mathcal{T}(c)Z^T = CZ^T \\ \Leftrightarrow \quad & C = U\Sigma(T^*)^{-1}(Z^T)^{-1} = U\Sigma\underbrace{(Z^TT^*)}_{T}{}^{-1}. \end{aligned}$$

Die Matrix T hat als Produkt regulärer Matrizen den vollen Rang und somit ist $k \in \mathcal{K}$.

□

Nun wird die Verallgemeinerung der Aussage von Satz 4 für nichtdiagonalisierbare Matrizen $M(k)$ vorbereitet. Der Lösungsraum eines Differentialgleichungssystems mit konstanter und diagonalisierbarer Koeffizientenmatrix besteht aus einer Menge von Exponentialfunktionen. Die Verallgemeinerung zu nichtdiagonalisierbaren Koeffizientenmatrizen führt auf die in Hilfssatz 5 definierte Menge von Funktionen. Zur Vorbereitung von Satz 5 wird nun die lineare Unabhängigkeit dieser Menge gezeigt.

Hilfssatz 5. *Seien $\lambda_i \neq \lambda_j$ für $i \neq j$ mit $i, j \leq n$ und $\mu_i \in \mathbb{N}$ mit $\mu_i > 0$. Dann sind die Funktionen der Menge $\{t^j\mathrm{e}^{\lambda_i t} : 0 \leq j \leq \mu_i, 1 \leq i \leq n\}$ linear unabhängig.*

Beweis: Es sei

$$f(t) := \sum_{i=1}^{n}\sum_{j=0}^{\mu_i}\alpha_{i,j}t^j\mathrm{e}^{\lambda_i t} = \sum_{i=1}^{n}\mathrm{e}^{\lambda_i t}\underbrace{\sum_{j=0}^{\mu_i}\alpha_{i,j}t^j}_{p_i(t)} = \sum_{i=1}^{n}\mathrm{e}^{\lambda_i t}p_i(t). \tag{4.11}$$

Nun wird angenommen, dass Polynome $p_i(t)$ mit $i = 1, \ldots, n$ und mindestens einem $p_i(t) \not\equiv 0$ sowie minimalem n existieren, sodass $f(t)$ die Nullfunktion ist. Mittels der

Ableitung des letzten Terms aus Gleichung (4.11) folgt

$$0 = f'(t) = \sum_{i=1}^{n} \lambda_i \mathrm{e}^{\lambda_i t} p_i(t) + \sum_{i=1}^{n} \mathrm{e}^{\lambda_i t} p_i'(t)$$

und mit $f(t) = f'(t) = 0$ gilt weiter

$$\begin{aligned}
0 &= (\lambda_n p_n(t) + p_n'(t)) f(t) - p_n(t) f'(t) \\
&= (\lambda_n p_n(t) + p_n'(t)) \sum_{i=1}^{n} \mathrm{e}^{\lambda_i t} p_i(t) - p_n(t) \left(\sum_{i=1}^{n} \lambda_i \mathrm{e}^{\lambda_i t} p_i(t) + \sum_{i=1}^{n} \mathrm{e}^{\lambda_i t} p_i'(t) \right) \\
&= (\lambda_n p_n(t) + p_n'(t)) \sum_{i=1}^{n-1} \mathrm{e}^{\lambda_i t} p_i(t) - p_n(t) \left(\sum_{i=1}^{n-1} \lambda_i \mathrm{e}^{\lambda_i t} p_i(t) + \sum_{i=1}^{n-1} \mathrm{e}^{\lambda_i t} p_i'(t) \right) \\
&= \sum_{i=1}^{n-1} \mathrm{e}^{\lambda_i t} \underbrace{\left[(\lambda_n p_n(t) + p_n'(t)) p_i(t) - \lambda_i p_n(t) p_i(t) - p_n(t) p_i'(t) \right]}_{q_i(t)} .
\end{aligned}$$

Da nun die Summe nur bis $n-1$ läuft und n minimal gewählt wurde, gilt $q_i(t) = 0$ für alle t und $i = 1, \ldots, n-1$. Es folgt für $i = 1, \ldots, n-1$

$$(\lambda_n - \lambda_i) p_n(t) p_i(t) = -p_n'(t) p_i(t) + p_n(t) p_i'(t) \ . \tag{4.12}$$

Wird nun angenommen, dass $n \geq 2$ gilt, folgt für den Term der linken Seite der Gleichung (4.12), dass der Grad $\deg(p_n) + \deg(p_i)$ ist, und für die rechte Seite ergibt sich ein maximaler Grad von $\deg(p_n) + \deg(p_i) - 1$. Dies ist ein Widerspruch, es folgt $n = 1$ und somit die Vereinfachung von Gleichung (4.11) zu

$$0 = \mathrm{e}^{\lambda_{i^*} t} p_{i^*}(t)$$

mit $i^* \in \{1, \ldots, n\}$. Weil nun aber aufgrund der Annahme $p_{i^*}(t)$ nicht die Nullfunktion ist, muss $\mathrm{e}^{\lambda_{i^*} t}$ die Nullfunktion sein. Somit ergibt sich der Widerspruch zur anfänglichen Annahme und es folgt $p_i(t) = 0$ für alle t und $i = 1, \ldots, n$ beziehungsweise $\alpha_{i,j} = 0$ für alle i, j. □

Für Satz 4 wurde unter anderem das trivialerweise gültige Kommutieren zweier Diagonalmatrizen benötigt. Dieser Zusammenhang muss in Vorbereitung von Satz 5 für Block-Toeplitz-Matrizen verallgemeinert werden.

Definition 2 (Persymmetrische Matrix, (obere) Toeplitz-Matrix)**.** Seien $A \in \mathbb{R}^{n \times n}$ und weiter $E \in \mathbb{R}^{n \times n}$ eine Antidiagonalmatrix mit $E_{i,n-i+1} = 1$ für $i = 1, \ldots, n$.
Eine Matrix A, für die $AE = EA^T$ gilt, heißt *persymmetrisch*. Eine persymmetrische Matrix ist also symmetrisch bezüglich der Antidiagonalen.
Gilt zusätzlich für die Elemente entlang einer jeden Neben- und der Hauptdiagonalen von A die Gleichheit, so heißt A *Toeplitz-Matrix*. Ist nur das obere Dreieck dieser Matrix A besetzt, heißt sie *obere Toeplitz-Matrix*.

Hilfssatz 6. *Die oberen Toeplitz-Matrizen $A, B \in \mathbb{R}^{n \times n}$ mit*

$$A := \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_2 \\ 0 & & & a_1 \end{pmatrix} \quad \textit{und} \quad B := \begin{pmatrix} b_1 & b_2 & \ldots & b_n \\ & \ddots & \ddots & \vdots \\ & & \ddots & b_2 \\ 0 & & & b_1 \end{pmatrix}$$

kommutieren.

Beweis: Es wird nun gezeigt, dass das Produkt AB ebenfalls eine obere Toeplitz-Matrix ist. Hieraus folgt im Anschluss unmittelbar, dass A und B kommutieren.
Weil das Produkt zweier oberer Dreiecksmatrizen wieder eine obere Dreiecksmatrix ist, braucht nur das obere Dreieck des Produkts AB betrachtet zu werden. Ein beliebiges Element des Produktes $(AB)_{i,j},\ i \leq j$, lässt sich wegen der Dreiecksform der Faktoren A und B als

$$(AB)_{i,j} = \sum_{l=i}^{j} a_{l-i+1} b_{j-l+1}$$

darstellen. Es gilt für $k < i \leq j$

$$\begin{aligned}(AB)_{i,j} &= \sum_{l=i}^{j} a_{l-i+1} b_{j-l+1} = \sum_{l=i}^{j} a_{(l-k)-(i-k)+1} b_{(j-k)-(l-k)+1} \\ &= \sum_{l=i-k}^{j-k} a_{l-(i-k)+1} b_{(j-k)-l+1} = (AB)_{i-k,j-k}\ .\end{aligned}$$

Zwei Elemente auf der Haupt- oder einer Nebendiagonalen des Produkts sind also identisch. Es folgt, dass das Produkt eine obere Toeplitz-Matrix und somit auch persymmetrisch ist. Aus der Persymmetrie folgt dann $ABE = E(AB)^T$ beziehungsweise $AB = E(AB)^T E^T$ mit E wie in Definition 2. Mit der Persymmetrie von A und B folgt

$$\begin{aligned}& E(AB)^T E = AB = AEEB = EA^T B^T E = E(BA)^T E \\ \Leftrightarrow\quad & AB = BA\end{aligned}$$

□

Bemerkung 2. Die Aussage von Hilfssatz 6 gilt auch, wenn eine Blockdiagonalstruktur für $A = \operatorname{diag}(A_1, \ldots, A_p)$ und $B = \operatorname{diag}(B_1, \ldots, B_p)$ mit Blöcken A_i und B_i gleicher Dimension für $i = 1, \ldots, p$ vorliegt.

Satz 5. *Die Aussage von Satz 4 gilt ebenfalls, wenn M und M^* nicht diagonalisierbar sind und die Vielfachheit der Eigenwerte berücksichtigt wird.*

Beweis: Die Beweisführung erfolgt analog zum Satz 4, wobei im Folgenden lediglich auf die Unterschiede näher eingegangen wird. Die Variablen sind analog bezeichnet. Weiter seien $\lambda_1, \ldots, \lambda_p$ die Eigenwerte von M mit Vielfachheiten $\mu_1, \ldots, \mu_p$, $J = \operatorname{diag}(J_1, \ldots, J_p)$ die Jordanmatrix von M mit Jordanblöcken J_i zu λ_i für $i = 1, \ldots, p$ und X die Matrix der entsprechenden Eigen- und Hauptvektoren, sodass $M = XJX^{-1}$ gilt. Sei $\alpha \in \mathbb{R}^n$ mit $\alpha = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)^T$ und $\tilde{\alpha}_i \in \mathbb{R}^{\mu_i}$ für $i = 1, \ldots, p$, sodass $c_0 = X\alpha$ gilt.
"⇒": Die Lösung des Differentialgleichungssystems (4.2) lässt sich darstellen als

$$c(t) = B\left(\mathrm{e}^{\lambda_1 t} \quad t\mathrm{e}^{\lambda_1 t} \quad t^2\mathrm{e}^{\lambda_1 t} \quad \cdots \quad t^{\mu_i - 1}\mathrm{e}^{\lambda_1 t} \quad \mathrm{e}^{\lambda_2 t} \quad \cdots \quad \mathrm{e}^{\lambda_s t}\right)^T \tag{4.13}$$

und ersetzt Gleichung (4.6). Die Funktionen in den Komponenten des Vektors der rechten Seite in Gleichung (4.13) sind nach Hilfssatz 5 linear unabhängig, womit auch das B regulär ist. Alle weiteren Implikationen erfolgen analog.
"⇐": Die Matrizen X und Y setzen sich für nicht diagonalisierbare Matrizen neben den

Eigenvektoren auch aus den entsprechenden Hauptvektoren zusammen. Die Gleichungen (4.8) und (4.9) werden ersetzt durch

$$X^{-1}M^*X = J,$$
$$YMY^{-1} = J.$$

Des Weiteren werden die Diagonalmatrizen D_1 und D_2 durch reguläre Blockdiagonalmatrizen R und R' ersetzt. Es ist $R = \mathrm{diag}(R_1, \ldots, R_p)$ mit oberen Toeplitz-Matrizen $R_i \in \mathbb{R}^{\mu_i \times \mu_i}$ für $i = 1, \ldots, p$. Die Matrix R' ist analog definiert. Weil J ebenfalls eine Blockmatrix bestehend aus oberen Toeplitz-Matrizen J_i mit $i = 1, \ldots, p$ ist, folgt mit Hilfssatz 6 und Bemerkung 2, dass J mit R und R' kommutiert. Die Beweisführung kann bis zur Gleichung (4.10) fortgesetzt werden, wobei $\tilde{X} = XR$, $\tilde{Y} = YR'$ und $Z = \tilde{X}\tilde{Y}$ gilt. Es muss nun noch gezeigt werden, dass die Einträge von R und R' so gewählt werden können, dass $Z^{-1}c_0 = c_0$ gilt. Auch hier wird die Darstellung $c_0 = \tilde{X}\alpha$ genutzt, wobei nach Hilfssatz 2 $(\tilde{\alpha}_i)_{\mu_i} \neq 0$ für $i = 1, \ldots, p$ gelten muss. Es folgt analog zu Satz 4

$$\begin{aligned} Z^{-1}c_0 = c_0 \quad &\Leftrightarrow \quad c_0 = Zc_0 \quad \Leftrightarrow \quad \tilde{X}\alpha = \tilde{X}\tilde{Y}\tilde{X}\alpha \\ &\Leftrightarrow \quad \alpha = R'YXR\alpha \quad \Leftrightarrow \quad (R')^{-1}\alpha = YXR\alpha \\ &\Leftrightarrow \quad \beta = YX\gamma \end{aligned}$$

mit $\beta := (R')^{-1}\alpha$ und $\gamma := R\alpha$. Wegen der Regularität von R und R' überträgt sich die bereits genannte Anforderung an α auch auf β und γ. Nach Hilfssatz 3 lassen sich solche β und γ finden. Die abschließenden Beweisschritte sind identisch zu Satz 4.

□

Es treten innerhalb einer Menge D-konsistenter Parameter $\mathcal{K}$ trivialerweise entweder ausschließlich diagonalisierbare oder nicht diagonalisierbare Koeffizientenmatrizen $M(k)$ für $k \in \mathcal{K}$ auf. Die Sätze 4 und 5 umfassen also alle relevanten Fälle. Die Menge $\mathcal{K}$ ist also unabhängig von der Diagonalisierbarkeit der entsprechenden Matrizen $M(k)$ durch die Eigenwerte von $M(k)$ unter Berücksichtigung der Vielfachheit charakterisiert. Da die Eigenwerte der Matrizen $M(k)$ für $k \in \mathcal{K}$ gleich bleiben, können zwei für die Berechnung der Menge $\mathcal{K}$ wichtigen Eigenschaften gefolgert werden.

Bemerkung 3. Da die Elemente des Parametervektors $k \in \mathbb{R}^q_+$ eines kinetischen Modells nichtnegativ sind und die Spur einer Matrix gleich der Summe ihrer Eigenwerte ist, folgt für beliebige kinetische Modelle erster Ordnung dieses Abschnitts mit den Sätzen 4 und 5, dass

$$\kappa := \sum_{i=1}^{q} k_i = -\mathrm{tr}(M(k)) = -\sum_{i=1}^{s} \lambda_i = \mathrm{const} \quad \forall k \in \mathcal{K}$$
$$\Rightarrow \quad 0 \leq k_j = \kappa - \sum_{\substack{i=1 \\ i \neq j}}^{q} k_i$$

gilt. Ein erster wichtiger Punkt ist, dass hiermit die zur Darstellung der Menge $\mathcal{K}$ nötigen Dimensionen um eins reduziert werden können. Diese Reduktion wird im Verlauf der Arbeit ohne weitere Erwähnung genutzt. Weiter gilt für ein beliebiges $j \in \{1, \ldots, q\}$, dass die Darstellung von $\mathcal{K}$ im Raum der verbliebenen $q-1$ Parameter k_i mit $i = 1, \ldots, q$ und $i \neq j$ durch die $q-1$ Halbebenen $k_i \geq 0$ sowie die Halbebene $\sum_{i=1, i\neq j}^{q} k_i \leq \kappa$ beschränkt ist.

In der abschließenden Bemerkung wird eine Funktion als Resultat der Sätze 4 und 5 definiert, mit der geprüft werden kann, ob eine Parametrierung eines kinetischen Modells D-konsistent ist.

Bemerkung. Sei $\tilde{\sigma}$ die Abbildung einer Matrix auf ihre, als Spaltenvektor angeordneten und aufsteigend bezüglich der Vektorindizes sortierten, Eigenwerte. Damit wird die Funktion

$$F_{\text{eig}}(k) := \frac{\|\tilde{\sigma}(M(k)) - \tilde{\sigma}(M(k^*))\|_2}{\|\tilde{\sigma}(M(k^*))\|_2} \tag{4.14}$$

definiert, mit der für ein bekanntes $k^* \in \mathcal{K}$ überprüft werden kann, ob auch $k \in \mathcal{K}$ erfüllt ist. Es gilt $k \in \mathcal{K}$ genau dann, wenn $F_{\text{eig}}(k) = 0$ ist.

4.1.1. Kinetiken erster Ordnung mit zeitabhängigem Vorfaktor

Es wird nun eine Verallgemeinerung der Sätze 4 und 5 für Kinetiken der Form

$$\dot{c}(t) = F(t)M(k)c(t), \quad c(t_0) = c_0. \tag{4.15}$$

durchgeführt. Die Funktion $F : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ sei stetig und es gilt weiter $\lim\limits_{t\to\infty} F(t) > 0$. Der Bezug zu praxisorientierten Anwendungen wird in der folgenden Bemerkung kurz erläutert.

Bemerkung. Kinetiken, wie in Gleichung (4.15), werden zur Modellierung von Konzentrationsverläufen in photochemischen Reaktionen in UV/Vis-spektroskopischen Untersuchungen genutzt. Die Funktion $F(t)$ entspricht dem sogenannte photokinetischen Faktor. Dieser kann mittels der Formel $F(t) \approx \tilde{F}(t) = \frac{1-10^{-w(t)}}{w(t)}$ approximiert werden [83]. Die Funktion $w(t) > 0$ beschreibt die Absorption einer bekannten Anregungswellenlänge zu einem Zeitpunkt t. Der kritische Fall $\lim\limits_{t\to\infty} \tilde{F}(t) = 0$ kann nur auftreten, wenn $\lim\limits_{t\to\infty} w(t) = \infty$ gilt. Dies ist aber in praktischen Anwendungen bereits durch Limitierung der maximal messbaren Intensität eines Spektrometers, also $w(t) \leq$ const, ausgeschlossen.

Die Aussagen der zentralen Sätze 4 und 5 können analog für Anfangswertprobleme der verallgemeinerten Form (4.15) bewiesen werden. Hierbei sind die Herleitung der analytischen Lösung von (4.15) und die Verallgemeinerung von Hilfssatz 5 nicht trivial. Dies wird in den folgenden Hilfssätzen 7 und 8 behandelt.

Hilfssatz 7. *Das Anfangswertproblem* (4.15) *besitzt die Lösung*

$$c(t) = \mathrm{e}^{M(k)\int_0^t F(\tau)\mathrm{d}\tau} c_0.$$

Beweis: Das Anfangswertproblem (4.15) kann durch Bilden der Jordan Normalform von $M(k)$, wie in (4.2) auch, auf das Lösen eines Anfangswertproblems mit einem Jordanblock $J \in \mathbb{R}^{\mu\times\mu}$ als Koeffizientenmatrix

$$\dot{d}(t) = F(t)Jd(t), \quad d(t_0) = d_0. \tag{4.16}$$

zurückgeführt werden. Es muss also lediglich gezeigt werden, dass das Anfangswertproblem (4.16) die Lösung

$$d(t) = \mathrm{e}^{J\int_0^t F(\tau)\mathrm{d}\tau} d_0$$

besitzt. Seien nun $\gamma(t) := \int_0^t F(\tau)\mathrm{d}\tau$ und $J = \lambda I + N$ eine additive Zerlegung von J mit Nebendiagonalanteil N, welcher nur aus Einsen auf der ersten oberen Nebendiagonalen besteht. Damit ist

$$d(t) = \mathrm{e}^{J\gamma(t)} d_0 = \mathrm{e}^{(\lambda I+N)\gamma(t)} d_0 = \mathrm{e}^{\lambda I\gamma(t)}\mathrm{e}^{N\gamma(t)} d_0 = \mathrm{e}^{\lambda\gamma(t)}\mathrm{e}^{N\gamma(t)} d_0 \,. \tag{4.17}$$

Mit der Nilpotenz der Matrix N ergibt sich

$$\begin{aligned}\frac{\partial}{\partial t}\mathrm{e}^{N\gamma(t)} &= \frac{\partial}{\partial t}\sum_{i=0}^{\infty}\frac{\gamma(t)^i}{i!}N^i = \frac{\partial}{\partial t}\sum_{i=0}^{\mu}\frac{\gamma(t)^i}{i!}N^i = \sum_{i=1}^{\mu}\frac{iF(t)\gamma(t)^{i-1}}{i!}N^i \\ &= F(t)N\sum_{i=1}^{\mu}\frac{\gamma(t)^{i-1}}{(i-1)!}N^{i-1} = F(t)N\sum_{i=0}^{\mu-1}\frac{\gamma(t)^i}{i!}N^i = F(t)N\mathrm{e}^{N\gamma(t)} \,.\end{aligned}$$

Mit (4.17) folgt damit weiter

$$\begin{aligned}\dot{d}(t) &= \lambda F(t)\mathrm{e}^{\lambda\gamma(t)}\mathrm{e}^{N\gamma(t)} d_0 + \mathrm{e}^{\lambda\gamma(t)} F(t) N\mathrm{e}^{N\gamma(t)} d_0 \\ &= F(t)(\lambda I + N)\mathrm{e}^{\lambda\gamma(t)}\mathrm{e}^{N\gamma(t)} d_0 = F(t)J\mathrm{e}^{J\gamma(t)} d_0 = F(t)Jd(t)\end{aligned}$$

und somit die Behauptung.

□

Hilfssatz 8. *Seien $\lambda_i \neq \lambda_j$ für $i \neq j$ mit $i, j \leq n$ und $\mu_i \in \mathbb{N}$ mit $\mu_i > 0$. Die Funktion $F : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ sei stetig mit $\lim_{t\to\infty} F(t) > 0$ und weiter ist $\gamma(t) := \int_0^t F(\tau)\mathrm{d}\tau$. Die Funktionen der Menge*

$$\{\gamma(t)^j \mathrm{e}^{\lambda_i\gamma(t)} : 0 \leq j \leq \mu_i, 1 \leq i \leq n\} \tag{4.18}$$

sind linear unabhängig.

Beweis: Aufgrund der Eigenschaften von $F(t)$ ist $\gamma(t)$ streng monoton steigend, stetig und surjektiv auf dem Intervall $[0, \infty)$. Nach [39] existiert eine Umkehrfunktion γ^{-1} auf dem Intervall $[0, \infty)$, sodass mit der Koordinatentransformation $t = \gamma^{-1}(\tau)$ die Gleichung $\gamma(\gamma^{-1}(\tau)) = \tau$ gilt. Statt der Menge (4.18) kann also alternativ auch die Menge

$$\{(\gamma(\gamma^{-1}(\tau)))^j \mathrm{e}^{\lambda_i\gamma(\gamma^{-1}(\tau))} : 0 \leq j \leq \mu_i, 1 \leq i \leq n\} = \{\tau^j \mathrm{e}^{\lambda_i\tau} : 0 \leq j \leq \mu_i, 1 \leq i \leq n\}$$

betrachtet werden. Die lineare Unabhängigkeit folgt nach Hilfssatz 5. □

4.1.2. Strukturanalyse

Im folgenden Abschnitt werden kinetische Modelle mit $s = 2$ und $s = 3$ Komponenten sowie weitere Spezialfälle bezüglich der Herleitung von Gleichungen zur Beschreibung der Menge D-konsistenter Parameter $\mathcal{K}$ untersucht. Hierzu werden die Ergebnisse

aus Abschnitt 4.1 auf ausgewählte kinetische Modelle angewendet und die verschiedenen Mengen $\mathcal{K}$ auf Grundlage der Eigenwerte der entsprechenden Koeffizientenmatrizen $M(k)$ definiert. Für ausgewählte kinetische Modell wird darüber hinaus eine analytische Darstellung der Mengen $\mathcal{K}^+$ hergeleitet.

Zur Veranschaulichung der in diesem Abschnitt behandelten Mengen $\mathcal{K}$ und $\mathcal{K}^+$ werden simulierte Matrizen D genutzt. Die zugehörigen Spalten des Faktors S basieren auf dem Modellproblem 1, wobei die i-te Spalte durch

$$\mathcal{S}_i(x) = 0.9 \cdot \mathrm{e}^{-\frac{x-i}{0.3}} + 0.1 \cdot \mathrm{e}^{-\frac{x-i}{10}}$$

definiert ist. Der Faktor C resultiert aus der Auswertung der jeweiligen Kinetik auf einem hinreichend feinen Zeitgitter T_G mit (in den Beispielen angegebenen) Anfangskonzentrationen c_0 und einem Geschwindigkeitsparameter $k^* \in \mathcal{K}$. Es gilt also $C = C^{\mathrm{dgl}}(k^*)$. Die jeweilige simulierte Matrix D ergibt sich dann durch CS^T.

Um die verschiedenen kinetischen Modelle innerhalb des Abschnitts voneinander abzugrenzen, werden zu Beginn eines jeden Unterabschnitts die Reaktionsgleichung und die Koeffizientenmatrix $M(k)$ des Differentialgleichungssystems des kinetischen Modells in einer Übersicht zusammengefasst.

Zweikomponentensysteme

Der einfachste Fall eines kinetischen Modells zwischen zwei Komponenten lautet

Irreversible Elementarreaktion

$$\mathcal{X} \xrightarrow{k_1} \mathcal{Y} \quad , \quad M(k) = \begin{pmatrix} -k_1 & 0 \\ k_1 & 0 \end{pmatrix}$$

Das entsprechende Anfangswertproblem

$$\dot{c}(t) = \begin{pmatrix} \dot{c}_{\mathcal{X}}(t) \\ \dot{c}_{\mathcal{Y}}(t) \end{pmatrix} = M(k) \begin{pmatrix} c_{\mathcal{X}}(t) \\ c_{\mathcal{Y}}(t) \end{pmatrix} = M(k)c(t), \quad c(0) = c_0 = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \tag{4.19}$$

hat die Lösung

$$\begin{pmatrix} c_{\mathcal{X}}(t) \\ c_{\mathcal{Y}}(t) \end{pmatrix} = \begin{pmatrix} \gamma_1 \mathrm{e}^{-k_1 t} \\ \gamma_1 + \gamma_2 - \gamma_1 \mathrm{e}^{-k_1 t} \end{pmatrix}.$$

Sei nun ein $k^* \in \mathcal{K}$ mit $k^* = k_1^*$ bekannt. Das Spektrum der Matrix $M(k^*)$ lautet $\{0, -k^*\}$ und es ist offensichtlich, dass kein $k \neq k^*$ existiert, sodass $\{0, -k^*\} = \{0, -k\}$ gilt. Die Menge $\mathcal{K} = \{k^*\}$ besteht also nur aus genau einem Element und das entsprechende nichtnegative Vollrangfaktorisierungsproblem mit Regularisierung durch diese Kinetik ist eindeutig lösbar. Diese erste Beispielkinetik stellt in sofern einen Spezialfall dar, dass alle weiteren untersuchten kinetischen Modelle erster Ordnung nicht in jedem Fall eine eindeutige Lösbarkeit der jeweiligen Matrixfaktorisierungsaufgabe bedingen.

Reversible Reaktion [113]

$$\mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \quad , \quad M(k) = \begin{pmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix} \tag{4.20}$$

Nun wird das soeben behandelte kinetische Modell um eine Rückreaktion erweitert. Das

Anfangswertproblem (4.19) mit $M(k)$ aus (4.20) hat mit $\kappa := k_1 + k_2$ die Lösung

$$\begin{pmatrix} c_{\mathcal{X}}(t) \\ c_{\mathcal{Y}}(t) \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} \gamma_1(k_1 \mathrm{e}^{-\kappa t} + k_2) + \gamma_2 k_2(1 - \mathrm{e}^{-\kappa t}) \\ \gamma_1 k_1(1 - \mathrm{e}^{-\kappa t}) + \gamma_2(k_2 \mathrm{e}^{-\kappa t} + k_1) \end{pmatrix}. \tag{4.21}$$

Man spricht hierbei von einer sogenannten Gleichgewichtsreaktion, da die Komponenten $\mathcal{X}$ und $\mathcal{Y}$ für $t \to \infty$ im Verhältnis $c_{\mathcal{X}}(t)/c_{\mathcal{Y}}(t) = k_2/k_1$ vorliegen. Das Spektrum der Matrix $M(k^*)$ zu gegebenem $k^* = (k_1^*, k_2^*) \in \mathcal{K}$ mit $\kappa^* = k_1^* + k_2^*$ lautet $\{0, -k_1^* - k_2^*\}$ beziehungsweise $\{0, -\kappa^*\}$. Damit es mit dem einer Matrix $M(k)$ übereinstimmt, muss für einen Geschwindigkeitsparameter $k = (k_1, k_2)$ also $k_1 + k_2 = \kappa^*$ gelten. Es folgt die Darstellung

$$\mathcal{K} = \{k \in \mathbb{R}_+^2 : \kappa^* = k_1 + k_2\}$$

für die Menge der D-konsistenten Parameter. Darüber hinaus kann leicht eine parametrisierte Darstellung hergeleitet werden:

$$\mathcal{K} = \left\{ \begin{pmatrix} \theta \\ \kappa^* - \theta \end{pmatrix} \in \mathbb{R}^2 : 0 < \theta < \kappa^* \right\}. \tag{4.22}$$

Bemerkung. Die Spezialfälle, in denen der Parameter θ die Werte 0 und κ^* annimmt, sind in Gleichung (4.22) aufgrund der Positivität der Geschwindigkeitsparameter ausgeschlossen. Es wird kurz erläutert, warum dies keine Relevanz für praktische Anwendungen hat. Ein Beispiel sei durch die Anfangswerte $(\gamma_1, \gamma_2)^T = (1, 0)^T$, ein $k^* > 0$ und $\theta = 0$ (beziehungsweise $k = (0, \kappa^*)$) gegeben. Damit ist die Lösung des Anfangswertproblems

$$\begin{pmatrix} c_{\mathcal{X}}(t) \\ c_{\mathcal{Y}}(t) \end{pmatrix} = \frac{1}{\kappa^*} \begin{pmatrix} (k_1 \mathrm{e}^{-\kappa^* t} + k_2) \\ k_1(1 - \mathrm{e}^{-\kappa^* t}) \end{pmatrix} = \frac{1}{\kappa^*} \begin{pmatrix} \kappa^* \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \forall t.$$

Diese ist also unabhängig von t und für k gilt außerdem $\mathrm{rg}(\mathcal{T}c(t)) = 1$, was aber den Bedingungen an einen D-konsistenten Parameter widerspricht. Aus chemischer Sicht sind diese Spezialfälle damit gleichbedeutend, dass ein Teil des Reaktionssystems "inaktiv" ist. Allgemein sind zwei Sonderfälle zu unterscheiden. Erstens, gilt bezüglich eines Geschwindigkeitsparameter k die Ungleichung $\mathrm{rg}(\mathcal{T}c(t)) < s$, so ist k in Kombination mit dem jeweiligen kinetischen Modell nicht zur Beschreibung der entsprechenden Konzentrationsverläufe geeignet. Ist für ein k mit $k_i = 0$ auch $\mathrm{rg}(\mathcal{T}c(t)) = s$ erfüllt, kann möglicherweise eine Vereinfachung des kinetischen Modells durch Streichen der i-ten Reaktion vorgenommen werden.

Es wird nun anhand des Reaktionssystems (4.20) erläutert, wie eine analytische Beschreibung der Menge $\mathcal{K}^+$ in Form einer parametrisierten Darstellung analog zu (4.22) hergeleitet werden kann. Das Ziel ist also alle θ für Gleichung (4.22) zu bestimmen, sodass der zu $k = (\theta, \kappa^* - \theta)^T \in \mathcal{K}$ gehörende Faktor S nichtnegativ ist.

Die Betrachtungen erfolgen zunächst für die Annahme kontinuierlicher Funktionen $c_{\mathcal{X}}(t)$, $c_{\mathcal{Y}}(t)$, $a_{\mathcal{X}}(x)$ und $a_{\mathcal{Y}}(x)$, welche dann im diskreten Fall in die Spalten von C und S übergehen. Seien $a_{\mathcal{X}} := a_{\mathcal{X}}(x)$ und $a_{\mathcal{Y}} := a_{\mathcal{Y}}(x)$ die dem Faktor S zugeordneten und $c_{\mathcal{X}} := c_{\mathcal{X}}(t)$ und $c_{\mathcal{Y}} := c_{\mathcal{Y}}(t)$ die dem Faktor C zugeordneten Funktionen. Auf die Angabe der Abhängigkeit dieser Funktionen vom Geschwindigkeitsparameter k wird verzichtet. Die aus dem bekannten Geschwindigkeitsparameter k^* resultierenden Funktionen lauten in analoger Weise $c_{\mathcal{X}}^*, c_{\mathcal{Y}}^*, a_{\mathcal{X}}^*$ und $a_{\mathcal{Y}}^*$. Es gilt der bilineare Zusammenhang

$$c_{\mathcal{X}}\ a_{\mathcal{X}} + c_{\mathcal{Y}}\ a_{\mathcal{Y}} = c_{\mathcal{X}}^*\ a_{\mathcal{X}}^* + c_{\mathcal{Y}}^*\ a_{\mathcal{Y}}^*\ . \tag{4.23}$$

Wegen der Gleichheit der Eigenwerte gilt $\kappa = k_1^* + k_2^* = k_1 + k_2$. Es sind nun drei Fälle zu unterscheiden. Zunächst werden die Anfangswerte $\gamma_1 > 0$ und $\gamma_2 = 0$ betrachtet. Einsetzen von (4.21) in (4.23) und Zusammenfassen resultiert in

$$0 = \frac{\gamma_1}{\kappa}(k_1 a_{\mathcal{X}} - k_1^* a_{\mathcal{X}}^* - k_1 a_{\mathcal{Y}} + k_1^* a_{\mathcal{Y}}^*)\mathrm{e}^{-\kappa t} + \frac{\gamma_1}{\kappa}(k_2 a_{\mathcal{X}} - k_2^* a_{\mathcal{X}}^* + k_1 a_{\mathcal{Y}} - k_1^* a_{\mathcal{Y}}^*).$$

Mit $\frac{\gamma_1}{\kappa} > 0$ und einem Koeffizientenvergleich folgen

$$0 = k_1 a_{\mathcal{X}} - k_1^* a_{\mathcal{X}}^* - k_1 a_{\mathcal{Y}} + k_1^* a_{\mathcal{Y}}^*, \tag{4.24}$$
$$0 = k_2 a_{\mathcal{X}} - k_2^* a_{\mathcal{X}}^* + k_1 a_{\mathcal{Y}} - k_1^* a_{\mathcal{Y}}^*. \tag{4.25}$$

Werden die Gleichungen (4.24) und (4.25) addiert und zusammengefasst, ergibt sich:

$$0 = \kappa a_{\mathcal{X}} - \kappa a_{\mathcal{X}}^* \quad \Rightarrow \quad a_{\mathcal{X}} = a_{\mathcal{X}}^* \,.$$

Somit ist die Nichtnegativität für $a_{\mathcal{X}}$ in jedem Fall gegeben. Es erfolgt die Substitution von $a_{\mathcal{X}}$ durch $a_{\mathcal{X}}^*$ in Gleichungen (4.24) und das Resultat wird nach $a_{\mathcal{Y}}$ umgestellt. Durch die Forderung nach Nichtnegativität von $a_{\mathcal{Y}}$ ergeben sich die Äquivalenzen

$$\begin{aligned} & 0 \leq a_{\mathcal{Y}} = \frac{1}{k_1}(a_{\mathcal{X}}^*(k_1 - k_1^*) + a_{\mathcal{Y}}^* k_1^*) \\ \Leftrightarrow \quad & 0 \leq a_{\mathcal{Y}} = a_{\mathcal{X}}^*(k_1 - k_1^*) + a_{\mathcal{Y}}^* k_1^* \\ \Leftrightarrow \quad & k_1 \geq k_1^* \frac{a_{\mathcal{X}}^* - a_{\mathcal{Y}}^*}{a_{\mathcal{X}}^*} \,. \end{aligned} \tag{4.26}$$

Für den diskreten Fall der Auswertung von $a_{\mathcal{X}}$ und $a_{\mathcal{Y}}$ bezüglich eines Gitters $X_G = (x_1, \ldots, x_m)$ führt (4.26) auf folgende Bedingung:

$$\theta = k_1 \geq k_1^* \max_{i=1,\ldots,m} \frac{a_{\mathcal{X}}^*(x_i) - a_{\mathcal{Y}}^*(x_i)}{a_{\mathcal{X}}^*(x_i)} \,. \tag{4.27}$$

Der zweite Fall mit $\gamma_1 = 0$ und $\gamma_2 > 0$ kann durch Vertauschung der Bezeichnungen k_1 und k_2 sowie $a_{\mathcal{X}}$ und $a_{\mathcal{Y}}$ auf den ersten Fall zurückgeführt werden. Auch für $\gamma_1 > 0$ und $\gamma_2 > 0$ kann das eben gezeigten Schema angewandt werden. Es wird sich deswegen auf die Angabe der hergeleiteten Bedingungen beschränkt. Für den kontinuierlichen Fall (analog zu Bedingung (4.26)) folgt die Nichtnegativität von $a_{\mathcal{X}}$ und $a_{\mathcal{Y}}$ aus der Bedingung

$$k_1 \begin{cases} \leq \gamma_2 \dfrac{a_{\mathcal{X}}^* k_2^* + a_{\mathcal{Y}}^* k_1^*}{\gamma_1 a_{\mathcal{X}}^* + \gamma_2 a_{\mathcal{Y}}^*} & \text{für } 0 > \kappa\gamma_2 - (\gamma_1 + \gamma_2)k_1 \\ \geq \gamma_2 \dfrac{a_{\mathcal{X}}^* k_2^* + a_{\mathcal{Y}}^* k_1^*}{\gamma_1 a_{\mathcal{X}}^* + \gamma_2 a_{\mathcal{Y}}^*} & \text{für } 0 < \kappa\gamma_2 - (\gamma_1 + \gamma_2)k_1 \end{cases}$$

und analog für den diskreten Fall (siehe Bedingung (4.27)) für den Spektrenfaktor S aus

$$\theta = k_1 \begin{cases} \leq \min\limits_{i=1,\ldots,m} \gamma_2 \dfrac{a_{\mathcal{X}}^*(x_i) k_2^* + a_{\mathcal{Y}}^*(x_i) k_1^*}{\gamma_1 a_{\mathcal{X}}^*(x_i) + \gamma_2 a_{\mathcal{Y}}^*(x_i)} & \text{für } 0 > \kappa\gamma_2 - (\gamma_1 + \gamma_2)k_1 \\ \geq \min\limits_{i=1,\ldots,m} \gamma_2 \dfrac{a_{\mathcal{X}}^*(x_i) k_2^* + a_{\mathcal{Y}}^*(x_i) k_1^*}{\gamma_1 a_{\mathcal{X}}^*(x_i) + \gamma_2 a_{\mathcal{Y}}^*(x_i)} & \text{für } 0 < \kappa\gamma_2 - (\gamma_1 + \gamma_2)k_1 \end{cases} .$$

Exemplarisch sind in Abbildung 4.1 die Mengen $\mathcal{K}$, $\mathcal{K}^+$ (links) und die Menge der zugehörigen Faktoren C (mittig) sowie Faktoren S (rechts) für ein Modellproblem dargestellt.

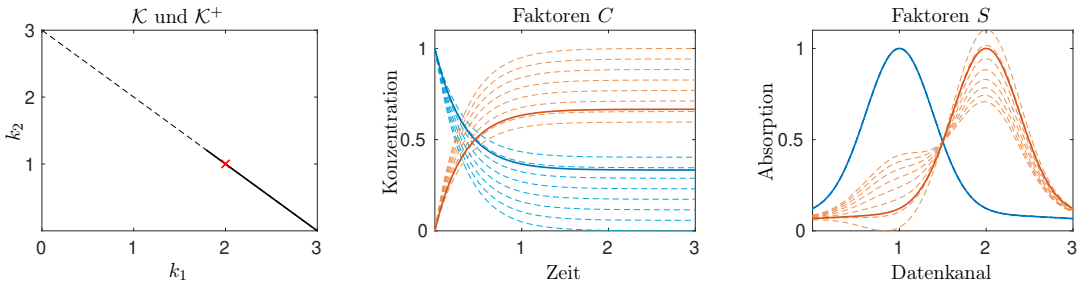
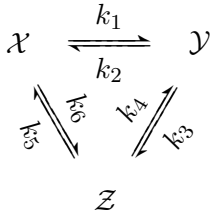


Abbildung 4.1.: Die linke Grafik zeigt eine Veranschaulichung der Mengen $\mathcal{K}$ (gestrichelt und durchgängig) und $\mathcal{K}^+$ (durchgängig) für das kinetische Modell $\mathcal{X} \leftrightarrow \mathcal{Y}$ mit $k^* = (k_1^*, k_2^*)^T = (2,1)^T$. Das rote Kreuz markiert k^*. In der Mitte sind Faktoren C und rechts Faktoren S für ausgewählte Elemente der Menge $\mathcal{K}^+$ dargestellt. Die Faktoren, welche zur Generierung des Modellproblems genutzt wurden, sind hierbei mit durchgängigen Linien gekennzeichnet.

Dreikomponentensysteme

Es werden nun kinetische Modelle betrachtet, an denen $s = 3$ Komponenten $\mathcal{X}, \mathcal{Y}$ und $\mathcal{Z}$ beteiligt sind. Ein sehr allgemeines Modell ist:



(4.28)

Das zugehörige Anfangswertproblem ist

$$\dot{c}(t) = \begin{pmatrix} \dot{c}_{\mathcal{X}}(t) \\ \dot{c}_{\mathcal{Y}}(t) \\ \dot{c}_{\mathcal{Z}}(t) \end{pmatrix} = \begin{pmatrix} -k_1 - k_6 & k_2 & k_5 \\ k_1 & -k_2 - k_3 & k_4 \\ k_6 & k_3 & -k_4 - k_5 \end{pmatrix} \begin{pmatrix} c_{\mathcal{X}}(t) \\ c_{\mathcal{Y}}(t) \\ c_{\mathcal{Z}}(t) \end{pmatrix} = M(k)c(t),$$
$$c(0) = c_0 = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}. \tag{4.29}$$

Es wird nun eine Darstellung der Eigenwerte von $M(k)$ hergeleitet. Sei A_i der i-te Minor der Matrix A, das heißt die Matrix, welche sich durch Streichen der i-ten Zeile und Spalte von A ergibt. Das charakteristische Polynom $cp(\lambda)$ der 3×3-Matrix $M := M(k)$ kann dargestellt werden als

$$\begin{aligned} cp(\lambda) &= \det(\lambda I - M) \\ &= \lambda^3 - \operatorname{tr}(M)\lambda^2 + (\det(M_1) + \det(M_3) + \det(M_3))\lambda + \det(M) \\ &= \lambda^3 - \operatorname{tr}(M)\lambda^2 + (\det(M_1) + \det(M_3) + \det(M_3))\lambda \\ &= \lambda(\underbrace{\lambda^2 - \operatorname{tr}(M)\lambda + \det(M_1) + \det(M_2) + \det(M_3)}_{p(\lambda)}). \end{aligned}$$

Eine Lösung von $cp(\lambda) = 0$ lautet offensichtlich $\lambda_1 = 0$ und es bleibt $p(\lambda) = 0$ zu betrachten. Mit $\kappa := \sum_{i=1}^{6} k_i = -\mathrm{tr}(M)$, $k = (k_1, \ldots, k_6)^T$ und

$$\alpha := \det(M_1) + \det(M_2) + \det(M_3) = \sum_{i<j} k_i k_j - \sum_{i+1=j} k_i k_j - k_1 k_6$$

folgt

$$\begin{aligned}
\lambda_{2,3} &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{\kappa^2 - 4\alpha} \\
&= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{\sum_{i=1}^{6} k_i^2 + 2\sum_{i<j} k_i k_j - 4\left(\sum_{i<j} k_i k_j - \sum_{i+1=j} k_i k_j - k_1 k_6\right)} \\
&= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{\sum_{i=1}^{6} k_i^2 - 2\sum_{i<j} k_i k_j + 4\sum_{i+1=j} k_i k_j + 4k_1 k_6} \\
&= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{\sum_{i=1}^{6} k_i^2 - 2\sum_{\substack{i<j \\ 2\le|i-j|\le 4}} k_i k_j + 2\sum_{i+1=j} k_i k_j + 2k_1 k_6} \\
&= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{\sum_{i=1}^{6} k_i^2 - \sum_{2\le|i-j|\le 4} k_i k_j + \sum_{|i-j|=1} k_i k_j + k_1 k_6 + k_6 k_1}
\end{aligned}$$

$$= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k^T N k} \quad \text{mit } N = \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}. \tag{4.30}$$

Wird nun nur ein Teil des kinetischen Modells (4.28) betrachtet, können die Eigenwerte der Matrix $M(k)$ des reduzierten Modells durch Streichen von Zeilen und Spalten in der Matrix N, Anpassen von κ und Umnummerierung der k_i leicht bestimmt werden.

Es werden nun ausgewählte Dreikomponentensysteme betrachtet und explizite Darstellungen der jeweiligen Mengen $\mathcal{K}$ hergeleitet. Die Bestimmung der Eigenwerte λ_2 und λ_3 erfolgt auf Basis eines jeweils bekannten Geschwindigkeitsparameters $k^* \in \mathcal{K}$ und wird in den folgenden Abschnitten nicht mehr explizit erwähnt.

Irreversible Folgereaktion [113]

$$\mathcal{X} \xrightarrow{k_1} \mathcal{Y} \xrightarrow{k_2} \mathcal{Z}\,, \quad M(k) = \begin{pmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & 0 \end{pmatrix}$$

Für diesen Fall lautet das Spektrum von $M(k)$ offensichtlich $\{0, -k_1, -k_2\}$. Es kann allerdings auch durch Streichen der Zeilen $2, 4, 5$ und 6 von N aus (4.30), der Umbenennung

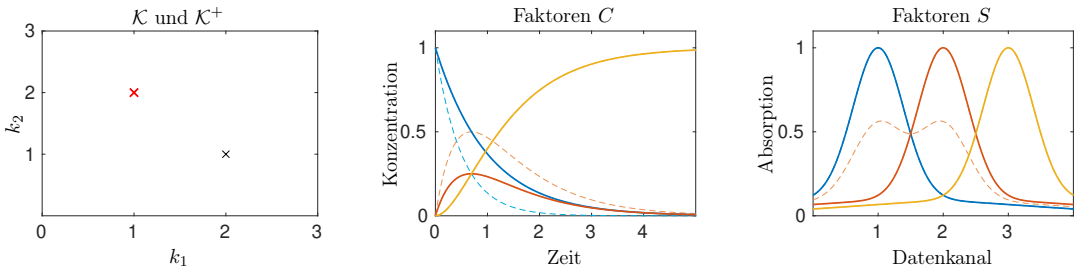


Abbildung 4.2.: In der linken Grafik sind die Mengen $\mathcal{K} = \mathcal{K}^+$ bestehend aus zwei Punkten für das kinetische Modell $\mathcal{X} \to \mathcal{Y} \to \mathcal{Z}$ mit $k^* = (k_1^*, k_2^*)^T = (1, 2)^T$ dargestellt. Das rote Kreuz markiert k^*. In der Mitte sind Faktoren C und rechts Faktoren S für ausgewählte Elemente der Menge $\mathcal{K}^+$ dargestellt. Die Faktoren, welche zur Generierung des Modellproblems genutzt wurden, sind hierbei mit durchgängigen Linien gekennzeichnet. Sie überdecken teilweise die Profile der zweiten Lösung.

von k_3 zu k_2 und $\kappa = k_1 + k_2$ hergeleitet werden:

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_{2,3} &= -\frac{k_1+k_2}{2} \pm \frac{1}{2}\sqrt{(k_1,k_2)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} k_1 \\ k_2 \end{pmatrix}} \\ &= -\frac{k_1+k_2}{2} \pm \frac{1}{2}\sqrt{(k_1-k_2)^2} = -\frac{k_1+k_2}{2} \pm \frac{k_1-k_2}{2} \\ \Rightarrow \quad \lambda_2 &= -k_2, \ \lambda_3 = -k_1 \, . \end{aligned}$$

Es ist leicht ersichtlich, dass die Forderung gleicher Eigenwerte von $M(k)$ auf die zwei Fälle $k_1 = -\lambda_3, k_2 = -\lambda_2$ und $k_1 = -\lambda_2, k_2 = -\lambda_3$ führt. Die Menge D-konsistenter Parameter lautet also

$$\mathcal{K} = \left\{ \begin{pmatrix} -\lambda_2 \\ -\lambda_3 \end{pmatrix}, \begin{pmatrix} -\lambda_3 \\ -\lambda_2 \end{pmatrix} \right\} .$$

Auf die Angabe der analytischen Beschreibung der Menge $\mathcal{K}^+$ wird verzichtet, da die beiden möglichen Parameterpaare in $\mathcal{K}$ ohne großen Aufwand bezüglich der Nichtnegativität der entsprechenden Faktoren S ausgewertet werden können. In Abbildung 4.2 sind die Ergebnisse eines Modellproblems zusammengefasst.

Die folgenden Kinetiken unterscheiden sich von den vorhergehenden dadurch, dass sie kaum bis gar nicht in der Literatur betrachtet und analysiert wurden.

Folgereaktion mit reversiblem ersten Teil [113]

$$\mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \xrightarrow{k_3} \mathcal{Z} \, , \quad M(k) = \begin{pmatrix} -k_1 & k_2 & 0 \\ k_1 & -k_2-k_3 & 0 \\ 0 & k_3 & 0 \end{pmatrix}$$

Durch Streichen der Zeilen und Spalten 4 bis 6 von N in (4.30) und mit $\kappa = k_1 + k_2 + k_3$

ergeben sich die Eigenwerte von $M(k)$ als

$$\begin{aligned}\lambda_1 &= 0,\\ \lambda_{2,3} &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k^T N_r k} \quad \text{mit } N_r = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}\\ &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k_1^2 + k_2^2 + k_3^2 + 2k_1k_2 + 2k_2k_3 - 2k_1k_3}\ .\end{aligned}$$

Auflösen von λ_2 nach k_3 und λ_3 nach k_2 ergibt

$$\begin{aligned}k_3 &= -\frac{\lambda_2(k_1 + k_2 + \lambda_2)}{k_1 + \lambda_2},\\ k_2 &= -\frac{k_1k_3 + k_1\lambda_3 + k_3\lambda_3 + \lambda_3^2}{\lambda_3}\end{aligned}$$

und es folgt durch Substitution

$$\begin{aligned}& k_3(k_1) := k_3 = \frac{\lambda_2\lambda_3}{k_1}\\ \Rightarrow \quad & k_2(k_1) := k_2 = \kappa - k_1 - k_3 = \kappa - k_1 - \frac{\lambda_2\lambda_3}{k_1} = -(\lambda_2 + \lambda_3) - k_1 - \frac{\lambda_2\lambda_3}{k_1}.\end{aligned}$$

Weiter ist $k_3(k_1) > 0$ für alle $k_1 > 0$ und es gilt mit

$$\begin{aligned}& 0 < k_2 = -\frac{k_1^2 + (\lambda_2 + \lambda_3)k_1 + \lambda_2\lambda_3}{k_1},\\ \Rightarrow \quad & 0 > k_1^2 + (\lambda_2 + \lambda_3)k_1 + \lambda_2\lambda_3 = (k_1 + \lambda_2)(k_1 + \lambda_3),\end{aligned}$$

dass $k_2(k_1) > 0$ für alle $k_1 \in (-\lambda_2, -\lambda_3)$. Die Menge $\mathcal{K}$ kann also für gegebene λ_2 und λ_3 als

$$\mathcal{K} = \left\{ \begin{pmatrix} k_1 \\ k_2(k_1) \\ k_3(k_1) \end{pmatrix} \in \mathbb{R}^3_+ : k_1 \in (-\lambda_2, -\lambda_3) \right\}$$

dargestellt werden. Analog zum Vorgehen für Zweikomponentensysteme können auch hier Grenzen für k_1 angegeben werden, welche auf einen nichtnegativen Faktor S führen. Im Folgenden wird das kinetische Modell nur für die Startkonzentrationen $c_0 = (1, 0, 0)^T$ betrachtet. Dies scheint eine starke Einschränkung zu sein, stellt aber in der praktischen Anwendung den wichtigsten Fall dar. Die Bezeichnung der Funktionen $a_\mathcal{X}, a_\mathcal{Y}$ und $a_\mathcal{Z}$ ist analog zum Abschnitt über Zweikomponentensysteme gewählt. Es folgen für den kontinuierlichen beziehungsweise diskreten Fall

$$k_1 \geq k_1^* \frac{a_\mathcal{X}^* - a_\mathcal{Y}^*}{a_\mathcal{X}^*} \quad \text{und} \quad k_1 \geq k_1^* \max_{i=1,\ldots,m} \frac{a_\mathcal{X}^*(x_i) - a_\mathcal{Y}^*(x_i)}{a_\mathcal{X}^*(x_i)} =: \alpha\ .$$

Hiermit kann auch die Menge zulässiger Parameter für den diskreten Fall

$$\mathcal{K}^+ = \left\{ \begin{pmatrix} k_1 \\ k_2(k_1) \\ k_3(k_1) \end{pmatrix} \in \mathbb{R}^3_+ : k_1 \in (\max(-\lambda_2, \alpha), -\lambda_3) \right\}$$

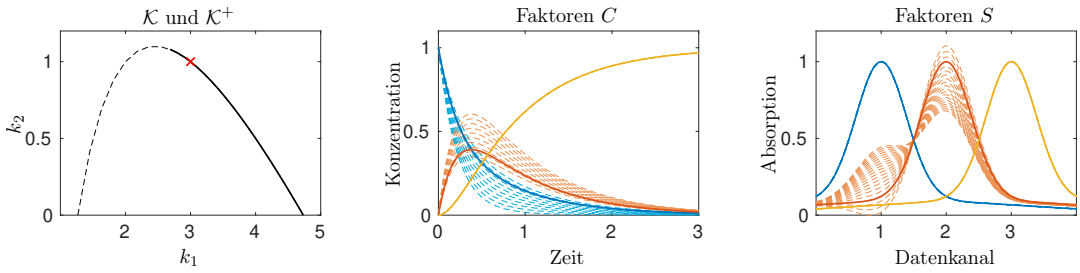


Abbildung 4.3.: In der linken Grafik sind die Mengen $\mathcal{K}$ (gestrichelt und durchgängig) und $\mathcal{K}^+$ (durchgängig) bestehend aus einer Kurve für das kinetische Modell $\mathcal{X} \leftrightarrow \mathcal{Y} \to \mathcal{Z}$ mit $k^* = (k_1^*, k_2^*, k_3^*)^T = (3, 1, 2)^T$ dargestellt. Das rote Kreuz markiert k^*. In der Mitte sind Faktoren C und rechts Faktoren S für ausgewählte Elemente der Menge $\mathcal{K}^+$ dargestellt. Die Faktoren, welche zur Generierung des Modellproblems genutzt wurden, sind hierbei mit durchgängigen Linien gekennzeichnet.

explizit angegeben werden.
In Abbildung 4.3 sind die Ergebnisse eines Modellproblems zusammengefasst.

Folgereaktion mit reversiblem zweiten Teil [113]

$$\mathcal{X} \xrightarrow{k_1} \mathcal{Y} \underset{k_3}{\overset{k_2}{\rightleftharpoons}} \mathcal{Z}\,, \qquad M(k) = \begin{pmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & k_3 \\ 0 & k_2 & -k_3 \end{pmatrix}$$

Analog zu den vorherigen Beispielen ergeben sich die Eigenwerte von $M(k)$ mit $\kappa = k_1 + k_2 + k_3$ als

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_{2,3} &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k^T N_r k} \quad \text{mit } N_r = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix} \\ &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k_1^2 + k_2^2 + k_3^2 + 2k_1k_2 + 2k_2k_3 - 2k_1k_3} \\ &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{(k_1 - k_2 - k_3)^2}, \\ \Rightarrow \quad \lambda_2 &= -k_1, \ \lambda_3 = -k_2 - k_3\,. \end{aligned}$$

Es folgt unmittelbar die Darstellung der Menge D-konsistenter Parameter

$$\mathcal{K} = \left\{ \begin{pmatrix} -\lambda_2 \\ k_2 \\ -\lambda_3 - k_2 \end{pmatrix} : k_2 \in (0, -\lambda_3) \right\} \cup \left\{ \begin{pmatrix} -\lambda_3 \\ k_2 \\ -\lambda_2 - k_2 \end{pmatrix} : k_2 \in (0, -\lambda_2) \right\}. \tag{4.31}$$

Auch für dieses kinetische Modell lässt sich die Menge $\mathcal{K}^+$ explizit angeben. Die beiden Teilmengen in (4.31) seien mit $\mathcal{K}_1$ und $\mathcal{K}_2$ bezeichnet. Ohne Beschränkung der Allgemeinheit enthalte $\mathcal{K}_1$ die bekannte Parametrierung k^*. Für $\mathcal{K}_1^+$ gilt dann

$$k_3 \leq \frac{k_3^* a_{\mathcal{Y}}^* + k_2^* a_{\mathcal{Z}}^*}{a_{\mathcal{Y}}^*} \quad \text{beziehungsweise} \quad k_3 \leq \min_{i=1,\dots,m} \frac{k_3^* a_{\mathcal{Y}}^*(x_i) + k_2^* a_{\mathcal{Z}}^*(x_i)}{a_{\mathcal{Y}}^*(x_i)} := \alpha.$$

Weiter kann gezeigt werden, dass $\mathcal{K}_2^+ \neq \emptyset$, wenn die Ungleichung $(\lambda_2 - \lambda_3)a_{\mathcal{X}}^* - \lambda_2 a_{\mathcal{Y}}^* > 0$

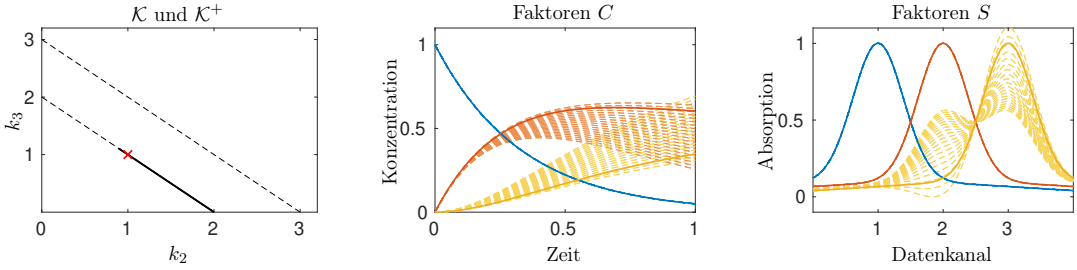


Abbildung 4.4.: In der linken Grafik sind die Mengen $\mathcal{K}$ (gestrichelt und durchgängig) und $\mathcal{K}^+$ (durchgängig) für das kinetische Modell $\mathcal{X} \to \mathcal{Y} \leftrightarrow \mathcal{Z}$ mit $k^* = (k_1^*, k_2^*, k_3^*)^T = (3,1,1)^T$ dargestellt. Das rote Kreuz markiert k^*. In der Mitte sind Faktoren C und rechts Faktoren S für ausgewählte Elemente der Menge $\mathcal{K}^+$ dargestellt. Die Faktoren, welche zur Generierung des Modellproblems genutzt wurden, sind hierbei mit durchgängigen Linien gekennzeichnet.

erfüllt ist. Dann impliziert die Forderung nach Nichtnegativität

$$k_3 \leq -\frac{\lambda_3(k_3^* a_{\mathcal{Y}}^* + k_2^* a_{\mathcal{Z}}^*)}{(\lambda_2 - \lambda_3)a_{\mathcal{X}}^* - \lambda_2 a_{\mathcal{Y}}^*} \quad \text{bzw.} \quad k_3 \leq \min_{i=1,\dots,m} -\frac{\lambda_3(k_3^* a_{\mathcal{Y}}^*(x_i) + k_2^* a_{\mathcal{Z}}^*(x_i))}{(\lambda_2 - \lambda_3)a_{\mathcal{X}}^*(x_i) - \lambda_2 a_{\mathcal{Y}}^*(x_i)} := \beta.$$

Somit lässt sich die Menge $\mathcal{K}^+$ darstellen als

$$\mathcal{K}^+ = \left\{ \begin{pmatrix} -\lambda_2 \\ -\lambda_3 - k_3 \\ k_3 \end{pmatrix} : k_3 \in (0, \alpha) \right\} \cup \left\{ \begin{pmatrix} -\lambda_3 \\ -\lambda_2 - k_3 \\ k_3 \end{pmatrix} : k_3 \in (0, \beta) \right\},$$

wobei die zweite Teilmenge gegebenenfalls leer ist.
In Abbildung 4.4 sind die Ergebnisse eines Modellproblems zusammengefasst.

Reversible Folgereaktion

$$\mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathcal{Z}\,, \quad M(k) = \begin{pmatrix} -k_1 & k_2 & 0 \\ k_1 & -k_2 - k_3 & k_4 \\ 0 & k_3 & -k_4 \end{pmatrix}$$

Dieses Reaktionssystem wird als erstes untersuchtes Beispiel eine Menge $\mathcal{K}$ in Form einer zweidimensionalen Mannigfaltigkeit aufweisen. Eine ausführliche Herleitung der folgenden Darstellung der Menge $\mathcal{K}$ ist in [111] zu finden. Die Eigenwerte von $M(k)$ mit $\kappa = k_1 + k_2 + k_3 + k_4$ sind

$$\lambda_1 = 0,$$

$$\lambda_{2,3} = -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k^T N_r k} \quad \text{mit } N_r = \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

$$= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k_1^2 + k_2^2 + k_3^2 + k_4^2 + 2(k_1k_2 + k_2k_3 + k_3k_4 - k_1k_3 - k_1k_4 - k_2k_4)},$$

wobei $\lambda_2 \geq \lambda_3$ gilt. Auf analoge Weise zur "Folgereaktion mit reversiblem ersten Teil" folgt mit

$$k_2(k_3, k_4) := -\frac{1}{k_3}(k_3 + k_4 + \lambda_2)(k_3 + k_4 + \lambda_3),$$
$$k_1(k_3, k_4) := \kappa - k_2(k_3, k_4) - k_3 - k_4$$

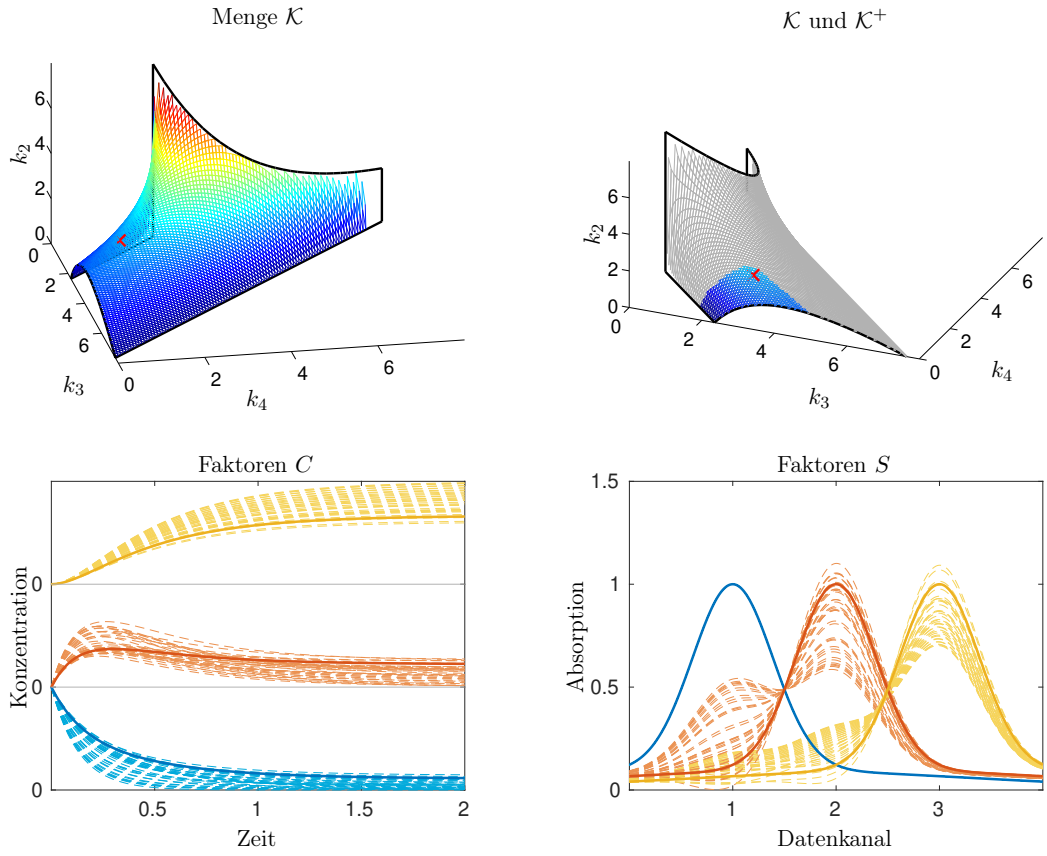


Abbildung 4.5.: In den oberen Grafiken sind links die Menge $\mathcal{K}$ und rechts eine Approximation der Menge $\mathcal{K}^+$ als farbige Teilmenge von $\mathcal{K}$ für das kinetische Modell $\mathcal{X} \leftrightarrow \mathcal{Y} \leftrightarrow \mathcal{Z}$ mit $k^* = (k_1^*, k_2^*, k_3^*, k_4^*)^T = (4, 2, 3, 1)^T$ dargestellt. Das rote Kreuz markiert k^*. Die schwarzen Linien zeigen die analytisch hergeleiteten Grenzen von $\mathcal{K}$. Die unteren Grafiken zeigen die Faktoren C und S für ausgewählte Elemente der Menge $\mathcal{K}^+$. Die Faktoren, welche zur Generierung des Modellproblems genutzt wurden, sind hierbei mit durchgängigen Linien gekennzeichnet.

die Menge D-konsistenter Parameter

$$\mathcal{K} = \left\{ \begin{pmatrix} k_1(k_3, k_4) \\ k_2(k_3, k_4) \\ k_3 \\ k_4 \end{pmatrix} : -\lambda_2 \leq k_3 + k_4 \leq -\lambda_3 \ \wedge \ k_3 > -\frac{1}{k_4}(k_4 + \lambda_2)(k_4 + \lambda_3) \right\}.$$

In Abbildung 4.5 sind die Ergebnisse eines Modellproblems zusammengefasst.

"Zyklisches Modell"

$$\mathcal{X} \xrightarrow{k_1} \mathcal{Y} \xrightarrow{k_2} \mathcal{Z} \xrightarrow{k_3} \mathcal{X}\,, \quad M(k) = \begin{pmatrix} -k_1 & 0 & k_3 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & -k_3 \end{pmatrix}$$

Mit $\kappa = k_1 + k_2 + k_3$ lauten die Eigenwerte von $M(k)$

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_{2,3} &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k^T N_r k} \quad \text{mit } N_r = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \\ &= -\frac{\kappa}{2} \pm \frac{1}{2}\sqrt{k_1^2 + k_2^2 + k_3^2 - 2k_1k_2 - 2k_2k_3 - 2k_1k_3}\,. \end{aligned}$$

Umformungen analog zur "Folgereaktion mit reversiblem ersten Teil" führen auf

$$k_2(k_1) := -\frac{1}{2}(k_1 + \lambda_2 + \lambda_3) \pm \frac{1}{2}\sqrt{-3k_1^2 + \lambda_2^2 + \lambda_3^2 - 2k_1(\lambda_2 + \lambda_3) - 2\lambda_2\lambda_3}\,.$$

Damit lässt sich die Menge D-konsistenter Parameter $\mathcal{K}$ durch eine Ellipsengleichung darstellen. Mit $E(k_1, k_2) := (k_2 - k_2(k_1))^2$ folgt

$$E(k_1,k_2) = \begin{pmatrix} k_1 & k_2 \end{pmatrix} \underbrace{\begin{pmatrix} -1 & -\frac{1}{2} \\ -\frac{1}{2} & -1 \end{pmatrix}}_{=:L} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} - (\lambda_2 + \lambda_3) \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} - \lambda_2\lambda_3 \tag{4.32}$$

und es gilt weiter

$$\mathcal{K} = \left\{ \begin{pmatrix} k_1 \\ k_2 \\ \kappa - k_1 - k_2 \end{pmatrix} \in \mathbb{R}^3_+ : E(k_1, k_2) = 0 \right\}. \tag{4.33}$$

Weiter wird nun noch die Parameterdarstellung von (4.33) bestimmt. Seien $\phi_1 = -1.5$ und $\phi_2 = -0.5$ die Eigenwerte von L, $\eta_1 = \eta_2 = -\frac{\lambda_2+\lambda_3}{3}$ die Koordinaten des Ellipsenmittelpunktes und

$$L_0 = \begin{pmatrix} -\lambda_2\lambda_3 & -\frac{\lambda_2+\lambda_3}{2} & -\frac{\lambda_2+\lambda_3}{2} \\ -\frac{\lambda_2+\lambda_3}{2} & -1 & -\frac{1}{2} \\ -\frac{\lambda_2+\lambda_3}{2} & -\frac{1}{2} & -1 \end{pmatrix}.$$

Die Längen der Halbachsen lauten

$$a = \sqrt{-\frac{|L_0|}{\phi_1|L|}} = \sqrt{\frac{1}{4.5}(\lambda_2^2 + \lambda_3^2 + \lambda_2\lambda_3)} \quad \text{und} \quad b = \sqrt{-\frac{|L_0|}{\phi_2|L|}} = \sqrt{\frac{1}{1.5}(\lambda_2^2 + \lambda_3^2 + \lambda_2\lambda_3)}\,.$$

Weiter ist die Ellipse um den Winkel $\tau = \frac{\pi}{4}$ rotiert. Mit

$$\begin{pmatrix} k_1(\theta) \\ k_2(\theta) \end{pmatrix} := \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \cos(\tau) & -\sin(\tau) \\ \sin(\tau) & \cos(\tau) \end{pmatrix} \begin{pmatrix} a\cos(\theta) \\ b\sin(\theta) \end{pmatrix} \tag{4.34}$$

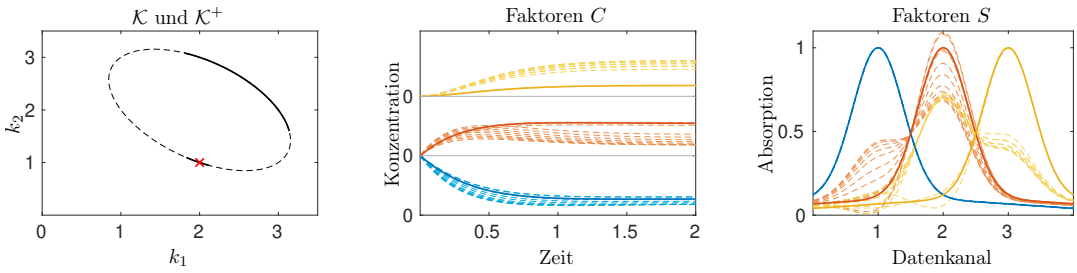


Abbildung 4.6.: In der linken Grafik sind die Mengen $\mathcal{K}$ (gestrichelt und durchgängig) und $\mathcal{K}^+$ (durchgängig) basierend auf einer Ellipse für das kinetische Modell $\mathcal{X} \to \mathcal{Y} \to \mathcal{Z} \to \mathcal{X}$ mit $k^* = (k_1^*, k_2^*, k_3^*)^T = (2, 1, 3)^T$ dargestellt. Das rote Kreuz markiert k^*. In der Mitte sind Faktoren C und rechts Faktoren S für ausgewählte Elemente der Menge $\mathcal{K}^+$ dargestellt. Die Faktoren, welche zur Generierung des Modellproblems genutzt wurden, sind hierbei mit durchgängigen Linien gekennzeichnet.

ergibt sich die Parameterdarstellung

$$\mathcal{K} = \left\{ \begin{pmatrix} k_1(\theta) \\ k_2(\theta) \\ \kappa - k_1(\theta) - k_2(\theta) \end{pmatrix} : \theta \in [0, 2\pi) \right\}.$$

In Abbildung 4.6 sind die Ergebnisse eines Modellproblems zusammengefasst.

Bemerkung. Unabhängig von der gewählten Kinetik wird die Menge $\mathcal{K}$ durch

$$k_i \geq 0 \ \forall i \quad \text{und} \quad \sum_{i=1, i \neq i^*}^{q} k_i \leq \kappa \ \forall i^* \tag{4.35}$$

beschränkt, siehe Bemerkung 3. Für das zyklische Modell können sowohl reelle als auch komplexe Eigenwerte λ_2 und λ_3 auftreten. Im Falle komplexer Eigenwerte sind die Ungleichungen (4.35) für die durch (4.32) beziehungsweise (4.34) definierte Ellipse alle erfüllt, das heißt $\mathcal{K}$ umfasst die vollständige Ellipse. Für den Fall reeller Eigenwerte setzt sich die Menge $\mathcal{K}$ nur noch aus Ellipsensegmenten zusammen.

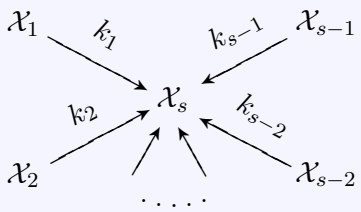
Spezialfälle

Allgemeine irreversible Folgereaktion

$$\mathcal{X}_1 \xrightarrow{k_1} \mathcal{X}_2 \xrightarrow{k_2} \dots \xrightarrow{k_{s-2}} \mathcal{X}_{s-1} \xrightarrow{k_{s-1}} \mathcal{X}_s \ ,$$

$$M(k) = \begin{pmatrix} -k_1 & & & 0 \\ k_1 & -k_2 & & \\ & k_2 & \ddots & \vdots \\ & & \ddots & -k_{s-1} \\ & & & k_{s-1} \quad 0 \end{pmatrix}$$

Irreversible “Sternreaktion”



$$M(k) = \begin{pmatrix} -k_1 & & & & 0 \\ & -k_2 & & & \\ & & \ddots & & \vdots \\ & & & -k_{s-1} & \\ k_1 & k_2 & \cdots & k_{s-1} & 0 \end{pmatrix}$$

Diese zwei kinetischen Modelle mit beliebiger Komponentenanzahl s umfassen mit der gegebenen Nummerierung der Komponenten keine Rückreaktionen, also Reaktionen von einem $\mathcal{X}_i$ zu einem $\mathcal{X}_j$ mit $i > j$. Dies hat zur Folge, dass oberhalb der Diagonalen von $M(k)$ nur Nulleinträge zu finden sind. Die Eigenwerte von $M(k)$ lauten dann

$$\lambda_1 = -k_1,\ \ldots\ ,\ \lambda_{s-1} = -k_{s-1},\ \lambda_s = 0.$$

Damit besitzt die Menge D-konsistenter Parameter für beide Kinetiken die folgende Darstellung:

$$\mathcal{K} = \left\{ -\begin{pmatrix} \lambda_{p_1} \\ \vdots \\ \lambda_{p_{s-1}} \end{pmatrix} : p \in \mathbb{S}_{s-1} \right\}$$

Hierbei ist $\mathbb{S}_{s-1}$ die symmetrische Gruppe.

4.1.3. Simultane Analyse mehrerer Datensätze

In praktischen Anwendungen ist es häufig möglich ein Experiment unter verschiedenen Bedingungen durchzuführen oder mittels verschiedener spektroskopischer Messverfahren zu vermessen. Dies eröffnet die Möglichkeit Serien von spektroskopischen Datensätzen nicht nur einzeln, sondern simultan zu analysieren. Im thematischen Umfeld der Reinkomponentenzerlegung finden solche theoretischen Ansätze [119, 121] vermehrt Anwendung [2, 27, 96].

Seien nun also Matrizen $D_1, \ldots, D_p$ gegeben, für die jeweils eine nichtnegative Vollrangfaktorisierung zu bestimmen ist. Es kann grundlegend zwischen zwei Fällen unterschieden werden:

Erstens, es wird angenommen, dass der Faktor C für alle Matrizen $D_1, \ldots, D_p$ übereinstimmt. Ein solcher Fall tritt beispielsweise auf, wenn für ein chemisches Reaktionssystem mehrere verschiedene Messverfahren zum Einsatz kommen. Der dann zu betrachtenden nichtnegativen Vollrangfaktorisierung liegt somit

$$\underbrace{\begin{pmatrix} D_1 & \cdots & D_p \end{pmatrix}}_{D} = C \underbrace{\begin{pmatrix} S_1^T & \cdots & S_p^T \end{pmatrix}}_{S^T} \tag{4.36}$$

zugrunde. Um die in Kapitel 3 beschriebenen Methoden anwenden zu können, ist lediglich ein Zusammenfügen der Matrix D aus den Datensätzen $D_1, \ldots, D_p$ wie in (4.36) nötig. Eventuell auftretende Inkonsistenzen bezüglich der Stützstellen entlang der jeweiligen Zeitachsen können durch Interpolation oder Trunkierung von $D_1, \ldots, D_p$ bezüglich des jeweiligen Zeitgitters behoben werden.

Zweitens, es wird angenommen, dass der Faktor S für alle Matrizen $D_1, \ldots, D_p$ übereinstimmt. Dieser Fall tritt insbesondere dann auf, wenn eine chemische Reaktion unter verschiedenen Reaktionsbedingungen wiederholt wird. Häufig haben Druck- oder Temperaturänderungen einen Einfluss auf die Geschwindigkeitsparameter oder aber die Startkonzentrationen, der an der Reaktion beteiligten Komponenten, werden variiert. Dies resultiert üblicherweise in unterschiedlichen zeitlichen Konzentrationsverläufen der beteiligten Komponenten. Die zu lösende nichtnegative Vollrangfaktorisierung basiert auf

$$\underbrace{\begin{pmatrix} D_1 \\ \vdots \\ D_p \end{pmatrix}}_{D} = \underbrace{\begin{pmatrix} C_1 \\ \vdots \\ C_p \end{pmatrix}}_{C} S^T.$$

Bei der Anwendung der Methoden aus Kapitel 3 sind ganz analog die, im ersten Fall genannten, Punkte zu beachten. Mögliche Inkonsistenzen sind hierbei entlang der Datenkanalachse zu beheben. Es kommt weiter hinzu, dass ein gegebenenfalls genutztes kinetisches Modell jeweils zu den verschiedenen Reaktionsbedingungen neu ausgewertet werden muss. Zu den Faktoren $C_1, \ldots, C_p$ werden also die Matrizen $C_1^{\mathrm{dgl}}(k), \ldots, C_p^{\mathrm{dgl}}(k)$ der numerischen Auswertung der jeweiligen Anfangswertprobleme bestimmt. Die Auswertung von F_{dgl} in der Zielfunktion (3.4) aus Abschnitt 3.1 erfolgt dann für jedes Paar C_i und $C_i^{\mathrm{dgl}}(k)$ mit $i = 1, \ldots, p$. Eine detaillierte Beschreibung der Vorgehensweise ist in [111] zu finden.

Die weiterführende Analyse mittels der Menge D-konsistenter Parameter aus Abschnitt 4.1 kann für beide Fälle ohne Einschränkung genutzt werden.

4.2. Kinetiken zweiter Ordnung

Kinetische Modelle zweiter Ordnung dienen unter anderem der Beschreibung bimolekularer Elementarreaktionen, also Reaktionen deren Reaktanten aus genau zwei Molekülen bestehen. Eine allgemeine Betrachtung der Menge D-konsistenter Parameter $\mathcal{K}$ für ein solches Modell ist typischerweise nicht möglich, da eine hierfür nötige analytische Lösung des zugrunde liegenden nichtlinearen Differentialgleichungssystems oft nicht zugänglich ist. Zur Lösung eines entsprechenden Anfangswertproblems kann üblicherweise nur auf numerische Lösungsverfahren zurückgegriffen werden. Nichtsdestotrotz lassen sich Aussagen über die Parameterlösungsmenge $\mathcal{K}$ treffen, die für eine Vielzahl von Reaktionssystemen relevant sind. Hierzu werden in Abschnitt 4.2.1 Reaktionssysteme mit pseudoerster Ordnung betrachtet. Für kinetische Modelle zweiter Ordnung liegt außerdem die Vermutung nahe, dass sie (als Regularisierung verwendet) die eindeutige Lösbarkeit des regularisierten Matrixfaktorisierungsproblems bedingen. Diese Vermutung wird exemplarisch für eine Kinetik in Abschnitt 4.2.2 bewiesen.

4.2.1. Kinetiken mit pseudoerster Ordnung

In diesem Abschnitt wird der Fall von kinetischen Modellen mit sogenannter pseudoerster Ordnung betrachtet. Es handelt sich dabei zwar um Modelle, die unter anderem bimolekulare Teilreaktionen abbilden, aber dennoch unter Zuhilfenahme geeigneter Annahmen mittels kinetischer Modelle erster Ordnung approximiert werden können. Diese Überlegungen können sogar für den allgemeinen Fall von r-molekularen Reaktionen angestellt

werden. Hierbei ist allerdings anzumerken, dass bereits trimolekulare Reaktionen äußerst selten auftreten und Reaktionen mit $r \geq 4$ praktisch keine Rolle spielen.
Sei ein kinetisches Modell r-ter Ordnung mit der Form

$$\mathcal{X}_1 + \ldots + X_r \xrightarrow{k_1} \mathcal{Y} \tag{4.37}$$

gegeben. Das zugehörige nichtlineare Differentialgleichungssystem lautet

$$\begin{aligned} \dot{c}_{\mathcal{X}_i}(t) &= -k_1 \prod_{j=1}^{r} c_{\mathcal{X}_j}(t), \quad i = 1, \ldots, r \\ \dot{c}_{\mathcal{Y}}(t) &= \quad k_1 \prod_{j=1}^{r} c_{\mathcal{X}_j}(t) \end{aligned}$$

Gilt im Vergleich zu $c_{\mathcal{X}_1}(t)$, dass die Funktionen $c_{\mathcal{X}_i} := c_{\mathcal{X}_i}(t),\ i = 2, \ldots, r$, näherungsweise konstant bezüglich t sind, kann (4.37) mit $\alpha = \prod_{i=2}^{r} c_{\mathcal{X}_i}$ zu

$$\mathcal{X}_1 \xrightarrow{\alpha k_1} \mathcal{Y}$$

vereinfacht werden. Können darüber hinaus alle Teilreaktionen mit einer Ordnung größer als 1 innerhalb eines Reaktionssystems so vereinfacht werden, hat das kinetische Modell eine sogenannte *pseudoerste* Ordnung. Die Ergebnisse aus Abschnitt 4.1 lassen sich dann analog übertragen.

4.2.2. Eindeutigkeit der nichtnegativen Matrixfaktorisierung für ausgewählte Kinetiken

Für kinetische Modelle zweiter Ordnung wurden bislang keine Aussagen zu damit regularisierten nichtnegativen Vollrangfaktorisierungen gemacht. Die Analyse ist im Vergleich zu Kinetiken erster Ordnung ungleich schwieriger, weil die Lösungen der entsprechenden Anfangswertprobleme häufig nicht analytisch bestimmt werden können. Für einfache Kinetiken können dennoch Aussagen über die Lösungsmengen von Geschwindigkeitsparametern hergeleitet werden.

Exemplarisch wird ein Reaktionsmodell der Form $2\mathcal{X} \to \mathcal{Y}$ mit dem zugehörigen Anfangswertproblem

$$\dot{c}(t) = \begin{pmatrix} \dot{c}_1(t) \\ \dot{c}_2(t) \end{pmatrix} = \begin{pmatrix} -2k_1 \\ k_1 \end{pmatrix} (c_1(t))^2, \quad c(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{4.38}$$

betrachtet. Im Folgenden wird gezeigt, dass eine durch die Kinetik (4.38) regularisierte Matrixfaktorisierung eine eindeutige Lösung besitzt.

Seien hierzu zwei Lösungen von (4.38) durch $c(t) = (c_1(t), c_2(t))^T$ zum Geschwindigkeitsparameter k_1 und $d(t) = (d_1(t), d_2(t))^T$ zum Geschwindigkeitsparameter $\tilde{k}_1$ gegeben. Es sei außerdem $d(t) = Yc(t)$ für eine reguläre Matrix $Y \in \mathbb{R}^{2\times 2}$. Weiter seien $c_1(t)$ und $c_2(t)$ linear unabhängig. Es wird nun gezeigt, dass Y die Einheitsmatrix sein muss. Für die beiden Lösungen gilt

$$\dot{c}(t) = \begin{pmatrix} -2k_1 \\ k_1 \end{pmatrix} c_1(t)^2, \quad \dot{d}(t) = \begin{pmatrix} -2\tilde{k}_1 \\ \tilde{k}_1 \end{pmatrix} d_1(t)^2 \quad \text{und} \quad c(0) = d(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Mit $d(t) = Yc(t)$ folgt

$$\dot{d}(t) = \begin{pmatrix} -2\tilde{k}_1 \\ \tilde{k}_1 \end{pmatrix} d_1(t)^2 \quad \Leftrightarrow \quad Y\dot{c}(t) = \begin{pmatrix} -2\tilde{k}_1 \\ \tilde{k}_1 \end{pmatrix} (Y_{1,1}c_1(t) + Y_{1,2}c_2(t))^2$$
$$\Leftrightarrow \quad Y \begin{pmatrix} -2k_1 \\ k_1 \end{pmatrix} c_1(t)^2 = \begin{pmatrix} -2\tilde{k}_1 \\ \tilde{k}_1 \end{pmatrix} (Y_{1,1}c_1(t) + Y_{1,2}c_2(t))^2 .$$

Zusammen mit der Regularität von Y und der linearen Unabhängigkeit von $c_1(t)$ und $c_2(t)$ ergibt sich durch einen Koeffizientenvergleich für $c_2(t)$, dass $Y_{1,2} = 0$ gilt. Für $t = 0$ folgt weiter, dass

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = d(0) = Yc(0) = Y \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} Y_{1,1} \\ Y_{2,1} \end{pmatrix}$$

gilt. Schlussendlich bleibt noch zu zeigen, dass $Y_{2,2} = 1$ ist. Hierzu wird $\dot{c}_1(t) + 2\dot{c}_2(t) = 0$ betrachtet und es folgt $c_1(t) + 2c_2(t) = c_1(0) + 2c_2(0) = 1$ für alle $t > 0$. Dies gilt analog auch für $d(t)$ und damit ist $1 = d_1(t) + 2d_2(t) = c_1(t) + 2Y_{2,2}c_2(t)$. Unter Berücksichtigung der Positivität von $c_1(t)$ und $c_2(t)$ für alle $t > 0$ nach Satz 3 folgt $Y_{2,2} = 1$ und somit auch die Behauptung $Y = I$.

4.3. Kritische Zusammenfassung

Die in Kapitel 3 eingeführten kinetischen Modelle können als Nebenbedingung effektiv zur Reduktion der Lösungsmenge des nichtnegativen Vollrangfaktorisierungsproblems einer nichtnegativen Matrix D eingesetzt werden. Diese Lösungsmengen, bestehend aus Faktorisierungen von D, lassen sich niedrigdimensional im Raum der Modellparameter durch die Mengen $\mathcal{K}$ und $\mathcal{K}^+$ darstellen. Die Parameterlösungsmengen von kinetischen Modellen erster und zweiter Ordnung werden in diesem Kapitel für störungsfreie Matrizen D analysiert.

Die in Abschnitt 4.1 auf Basis der Eigenwerte der Koeffizientenmatrix $M(k)$ eingeführte äquivalente Darstellung von $\mathcal{K}$ für Kinetiken mit erster Ordnung ist ein zentrales Ergebnis dieser Arbeit. Darauf aufbauend können zum Beispiel allgemeinere Klassen von Kinetiken (Abschnitt 4.1.1), die Struktur von $\mathcal{K}$ (Abschnitt 4.1.2) und die simultane Analyse mehrerer Matrizen D (Abschnitt 4.1.3) untersucht werden. Wie bei der Menge zulässiger Lösungen zur Darstellung nichtnegativer Matrixfaktorisierungen wird auch die Anwendung von $\mathcal{K}$ und $\mathcal{K}^+$ durch die gegebenenfalls notwendige grafische Darstellung limitiert. Durch Anwendung der Dimensionsreduktion aus Bemerkung 3 ist eine Darstellung von Kinetiken mit bis zu $q = 4$ Teilreaktionen uneingeschränkt möglich. Für $q \geq 5$ können entsprechende Projektionen genutzt werden.

Die in Abschnitt 4.2 betrachteten kinetischen Modelle zweiter Ordnung werden nicht in dieser Allgemeinheit untersucht, weil die Lösungen der zugrunde liegenden Anfangswertprobleme im Regelfall nicht analytisch zugänglich sind. Es wird sich daher auf den Spezialfall pseudoerster Ordnungen (Abschnitt 4.2.1) und den Nachweis der eindeutigen Lösbarkeit der regularisierten Matrixfaktorisierung für die Kinetik $2\mathcal{X} \to \mathcal{Y}$ (Abschnitt 4.2.2) beschränkt. Eine wichtige offene Frage ist, ob sich der letztgenannte Eindeutigkeitsnachweis auch für allgemeine kinetische Modelle zweiter Ordnung (unter Vernachlässigung von trivialen Fällen) durchführen lässt.

5. Störungsanalyse

In Abschnitt 4.1 des vorherigen Kapitels wurden die Begriffe der D-konsistenten und zulässigen Parameter zur Beschreibung von Parameterlösungsmengen kinetischer Modelle im Kontext nichtnegativer Vollrangfaktorisierungen eingeführt. Die Theorie basiert auf einer störungsfreien nichtnegativen Matrix D. Für experimentell ermittelte Ausgangsdaten ist D typischerweise störungsbehaftet. Entsprechende Verallgemeinerungen der Mengen D-konsistenter Parameter $\mathcal{K}$ und zulässiger Parameter $\mathcal{K}^+$ werden in Abschnitt 5.1 eingeführt. Inwiefern die Menge $\mathcal{K}$ auch im gestörten Fall als Approximation verallgemeinerten Parameterlösungsmengen einsetzbar ist, wird anhand einer entsprechenden Fehleranalyse untersucht. In Abschnitt 5.2 wird in Form einer Fallstudie die Michaelis-Menten-Kinetik bezüglich der Existenz nichttrivialer Parameterlösungsmengen unter dem Einfluss von Störungen untersucht. Abschließend werden in Abschnitt 5.3 sogenannte L-Kurven betrachtet, durch die sich geeignete Gewichtungen der Zielfunktionen der vorherigen Kapitel bestimmen lassen.

Einleitend wird darauf eingegangen, welche Störungen in einer experimentell gemessenen Spektrenfolge zu erwarten sind. Die folgende Bemerkung fasst drei Typen von Störungen zusammen, welche die Bestimmung einer approximativen/regularisierten Matrixfaktorisierung entscheidend beeinflussen.

Bemerkung. Typische Störungen spektroskopischer Messdaten und wodurch sie hervorgerufen werden, sind:

1. Rauschen: Durch kleine, zufällige Messungenauigkeiten hervorgerufen, zeichnen sich experimentell ermittelte Spektren im Regelfall durch geringe Schwankungen aus. Deren Intensität wird durch das Signal-Rausch-Verhältnis charakterisiert. Je nachdem, ob eine Abhängigkeit von der Signalintensität besteht oder nicht, wird zwischen hetero- und homoskedastischem Rauschen unterschieden [72].
2. Negative Einträge in Spektren: Ein Spektrum, welches negative Einträge besitzt, ist aus physikalischer Sicht, für die in dieser Arbeit untersuchten spektroskopischen Methoden (UV/Vis, IR, Raman), nicht möglich. Dennoch tritt diese Form der Störung nicht selten auf, wenn zum Beispiel das Spektrum eines Lösungsmittels zeilenweise von einer Spektrenserie $D \in \mathbb{R}^{m \times n}$ abgezogen wird.
3. Grundlinienstörungen: Das Auftreten von Nebenreaktionen, die nur einen kleinen additiven Anteil zu den Spektren einer Folge beitragen, können typischerweise als Störung betrachtet werden.

5.1. D-Konsistenz unter Berücksichtigung von Störungen

Die Mengen $\mathcal{K}$ und $\mathcal{K}^+$ sind für störungsfreie Ausgangsdaten D eingeführt. Sie lassen sich auch zur Analyse von experimentell ermittelten Spektrenfolgen, wie beispielsweise für den spektroskopische Datensatz 1 in Abschnitt 3.4, einsetzen. Die Analyse erfolgt dann unter der Annahme, dass keine Störungen vorliegen. Dies ist für geringe Störungen

vertretbar und die Ergebnisse sind zufriedenstellend. Dennoch ist klar, dass eine bessere Art der Handhabung von auftretenden Störungen wünschenswert ist. Entsprechende Verallgemeinerungen werden nun präsentiert.

Ein typisches Vorgehen besteht in der Einführung von Fehlertoleranzen, welche sich hier auf die Konsistenz des Faktors C mit der jeweiligen Modellauswertung $C^{\mathrm{dgl}}(k)$ und die Nichtnegativität des Faktors S beziehen werden. Es wird mit einer Verallgemeinerung der Menge $\mathcal{K}$ begonnen.

Definition 3 (Menge D-approximativer Parameter)**.** Seien die Matrix $D \in \mathbb{R}^{m\times n}$ mit $\mathrm{rg}(D) \geq s \geq 1$ und die abgeschnittene Singulärwertzerlegung $D \approx D_s = U\Sigma V^T$ gegeben. Für einen gegebenen Parametervektor $k \in \mathbb{R}_+^q$ ist der im Sinne kleinster Fehlerquadrate optimale Faktor $C(k) := (U\Sigma)(T(k))^+$ mit der Transformationsmatrix $T(k) := C^{\mathrm{dgl}}(k)^+(U\Sigma)$ definiert. Weiter sei eine Diagonalmatrix $B \in \mathbb{R}^{s\times s}$ mit positiven Diagonaleinträgen gegeben. Die Menge der *D-approximativen Parameter* ist durch

$$\mathcal{K}_\varepsilon := \left\{k \in \mathbb{R}^q : \mathrm{rg}(T(k)) = s \wedge \|(C(k) - C^{\mathrm{dgl}}(k))B^{-1}\|_F \leq \sqrt{ms}\ \cdot\varepsilon\right\}$$

definiert.

Es wird also gegenüber dem störungsfreien Fall $\mathcal{K}$ eine Toleranz ε bezüglich der Konsistenz von Kinetik und Faktor C zugelassen. Die Matrix B kann zur Skalierung der Spalten von C und $C^{\mathrm{dgl}}(k)$ genutzt werden, wenn diese stark unterschiedliche Größenordnungen aufweisen. Eine mögliche Wahl für ein bekanntes $k^* \in \mathcal{K}_\varepsilon$ ist beispielsweise

$$B = \mathrm{diag}\left(\max_l\left(C(k^*)_{l,1}\right), \ldots, \max_l\left(C(k^*)_{l,s}\right)\right).$$

Soll auf eine Skalierung verzichtet werden, ist $B = I$ zu verwenden. Des Weiteren wird der Skalierungsfaktor $\sqrt{ms}$ genutzt, um einen Einfluss der Anzahl der Zeitgitterstützstellen sowie der Anzahl der Komponenten auf die Wahl von ε weitestgehend zu reduzieren. Zur numerischen Berechnung eines Elements aus $\mathcal{K}_\varepsilon$ kann die Zielfunktion

$$F_\varepsilon(k) := \gamma_1\|I - T(k)T(k)^+\|_F^2 + \delta_1 \max\left(\|(C(k) - C^{\mathrm{dgl}}(k))B^{-1}\|_F - \sqrt{ms}\ \cdot\varepsilon, 0\right) \quad (5.1)$$

mit den Gewichtungsfaktoren γ_1 und δ_1 minimiert werden.

Bemerkung. Es handelt sich bei $\mathcal{K}_\varepsilon$ lediglich um *einen* möglichen Ansatz zur Untersuchung von gestörten Matrizen D. Eine Alternative ist beispielsweise

$$\mathcal{K}'_\varepsilon = \left\{k \in \mathbb{R}^q : \mathrm{rg}(T(k)) = s \wedge \left|\frac{(C(k))_{i,j} - (C^{\mathrm{dgl}}(k))_{i,j}}{B_{j,j}}\right| - \varepsilon \leq 0\ \forall i,j\right\},$$

wobei hier die Fehlertoleranz komponentenweise angewendet wird. Es handelt sich nicht um eine äquivalente Definition zu $\mathcal{K}_\varepsilon$, sondern es gilt lediglich die Beziehung $\mathcal{K}'_\varepsilon \subseteq \mathcal{K}_\varepsilon$.

Analog zur Einschränkung von $\mathcal{K}$ auf $\mathcal{K}^+$ unter Zuhilfenahme der Nichtnegativität der Faktoren S kann auch für $\mathcal{K}_\varepsilon$ eine entsprechende Teilmenge definiert werden. Hierzu werden betragskleine negative Einträge in S zugelassen.

Definition 4 (Menge zulässiger D-approximativer Parameter)**.** Seien die Voraussetzungen wie in Definition 3 und weiter $S(k) := VT(k)^T$. Die *Menge der zulässigen D-approximativen Parameter* ist mit $\theta \geq 0$ durch

$$\mathcal{K}^+_{\varepsilon,\theta} := \left\{k \in \mathcal{K}_\varepsilon : \frac{(S(k))_{i,j}}{\max_l(|(S(k))_{l,j}|)} + \theta \geq 0\ \forall i,j\right\}$$

definiert.

Der Parameter θ entspricht der prozentual zugelassenen komponentenweisen Negativität in Bezug auf die auf ein betragsmäßiges Maximum von 1 skalierten Spalten des Faktors S. Die numerische Berechnung eines Elements aus der Menge $\mathcal{K}^+_{\varepsilon,\theta}$ kann durch Minimierung der, analog zu (5.1) aufgestellten, Zielfunktion

$$F_{\varepsilon,\theta}(k) := F_\varepsilon(k) + \gamma_2 \min\left(\frac{(S(k))_{i,j}}{\max_l(|(S(k))_{l,j}|)} + \theta, 0\right) \tag{5.2}$$

mit Gewichtungsfaktor γ_2 durchgeführt werden. Es ist leicht nachzuweisen, dass für störungsfreie Modellprobleme D die Mengenbeziehungen

$$\mathcal{K} \subseteq \mathcal{K}_\varepsilon \text{ mit } \varepsilon \geq 0 \quad \text{und} \quad \mathcal{K}^+ \subseteq \mathcal{K}^+_{\varepsilon,\theta} \text{ mit } \varepsilon, \theta \geq 0$$

erfüllt sind. Diese stellen die Grundlage für eine mögliche Wahl der Parameter ε und θ dar. Die Idee besteht darin eine störungsbehaftete Matrix D als störungsfrei anzusehen und die Mengen $\mathcal{K}$ sowie $\mathcal{K}^+$ zu bestimmen. Anschließend ist das minimale ε zu bestimmen, sodass $\mathcal{K} \subseteq \mathcal{K}_\varepsilon$ erfüllt ist. Auf analoge Weise wird ein minimales θ bestimmt, für das zusätzlich die Mengenbeziehung $\mathcal{K}^+ \subseteq \mathcal{K}^+_{\varepsilon,\theta}$ gilt.

Die Resultate aus Abschnitt 4.1 zur Berechnung der Menge $\mathcal{K}$ mittels der Eigenwerte der Koeffizientenmatrix $M(k)$ des kinetischen Modells können nicht zur direkten Berechnung der Menge $\mathcal{K}_\varepsilon$ beziehungsweise $\mathcal{K}^+_{\varepsilon,\theta}$ eingesetzt werden. Der Grund liegt darin, dass für störungsbehaftete Matrizen D typischerweise keine D-konsistenten sondern nur D-approximative Parameter mit einem $\varepsilon > 0$ existieren. Die Voraussetzungen der Sätze 4 und 5 sind somit im Allgemeinen nicht erfüllt. Es muss auf die aufwendiger auszuwertenden Zielfunktionen (5.1) oder (5.2) zurückgegriffen werden, die die Lösung eines Anfangswertproblems beinhalten.

Eine Fehleranalyse der Approximation von $\mathcal{K}_\varepsilon$ durch $\mathcal{K}$

Soll lediglich eine Approximation von $\mathcal{K}_\varepsilon$ bestimmt werden, stellt sich die Frage, ob $\mathcal{K}$ hierfür ausreichend ist. Diese Betrachtung lohnt sich, weil die Berechnung von $\mathcal{K}_\varepsilon$ im Vergleich zu $\mathcal{K}$ mit einem erheblich größeren Rechenaufwand verbunden ist. Weil die Berechnung der Menge $\mathcal{K}$ nach Kapitel 4 nur von einem bekannten k^* abhängt, nicht aber von der Matrix D, kann $\mathcal{K}$ auch dann bestimmt werden, wenn k^* lediglich D-approximativ und nicht D-konsistent ist. Es wird hierzu die Notation $\mathcal{K}(k^*)$ eingeführt, um eine Abgrenzung zur Menge D-konsistenter Parameter zu erhalten. Die Berechnungen von $\mathcal{K}$ und $\mathcal{K}(k^*)$ sind also identisch, jedoch beschreibt $\mathcal{K}(k^*)$ im Kontext dieses Kapitels eine Menge D-approximativer Parameter. Es wird nun der maximale Fehler

$$\delta_{\max} = \max_{k \in \mathcal{K}(k^*)} \|(C(k) - C^{\mathrm{dgl}}(k))B^{-1}\|_F \tag{5.3}$$

der Approximation von $\mathcal{K}_\varepsilon$ durch $\mathcal{K}$ abgeschätzt. Für einen gegebenen D-approximativen Parameter k^* kann auch der Fehler der entsprechenden Kinetikanpassung

$$\delta = \|(C(k^*) - C^{\mathrm{dgl}}(k^*))B^{-1}\|_F$$

berechnet werden. Mit dem Satz 6 sowie den Folgerungen 1 und 2 wird im Folgenden eine Abschätzung für $\delta_{\max}$ auf Grundlage von δ hergeleitet.

Im folgenden Satz werden Aussagen über die Transformation zwischen den Matrizen $C^{\mathrm{dgl}}(k)$ und $C^{\mathrm{dgl}}(k^*)$ mit $k, k^* \in \mathcal{K}(k^*)$ sowie die zugehörigen Fehler getroffen.

Satz 6. *Seien die Voraussetzungen wie in Definition 3 und ein k^* gegeben. Für alle $k \in \mathcal{K}(k^*)$ gilt die Bezeichnung $T_k := T(k)$ und es seien die Residuen E_k und $\tilde{E}_k$ durch*

$$C(k) - C^{\mathrm{dgl}}(k) = U\Sigma T_k^+ - C^{\mathrm{dgl}}(k) = E_k = \tilde{E}_k T_k^+ \tag{5.4}$$

definiert. Dann gilt:

a) $\mathrm{Im}(U\Sigma) \perp \mathrm{Im}(E_k)$ *und* $\mathrm{Im}(U\Sigma) \perp \mathrm{Im}(\tilde{E}_k)$,

b) $\tilde{E}_{k^*} = \tilde{E}_k$,

c) $C^{\mathrm{dgl}}(k^*) = C^{\mathrm{dgl}}(k) T_k T_{k^*}^+$.

Beweis: Für die Fehlermatrix E_k bezüglich eines beliebigen $k \in \mathcal{K}(k^*)$ gilt

$$E_k = U\Sigma T_k^+ - C^{\mathrm{dgl}}(k) = U\Sigma (U\Sigma)^+ C^{\mathrm{dgl}}(k) - C^{\mathrm{dgl}}(k) = \underbrace{[(U\Sigma)(U\Sigma)^+ - I]}_{\text{Projektion in } \ker((U\Sigma)^T)} C^{\mathrm{dgl}}(k)$$

womit gezeigt ist, dass $0 = (U\Sigma)^T E_k$ und damit auch $0 = (U\Sigma)^T E_k = (U\Sigma)^T \tilde{E}_k T_k^+$, also die Behauptung a), gilt. Weiter folgt aus Gleichung (5.4) für alle $k \in \mathcal{K}(k^*)$

$$C^{\mathrm{dgl}}(k) = (U\Sigma - \tilde{E}_k) T_k^+. \tag{5.5}$$

Aus $k^*, k \in \mathcal{K}(k^*)$ folgt weiter, dass eine lineare Transformation $W \in \mathbb{R}^{s \times s}$ existiert, sodass

$$C^{\mathrm{dgl}}(k^*) = C^{\mathrm{dgl}}(k) W$$

gilt und mit Gleichung (5.5) folgt

$$\begin{aligned}
&(U\Sigma - \tilde{E}_{k^*}) T_{k^*}^+ = (U\Sigma - \tilde{E}_k) T_k^+ W \\
\Rightarrow\ & U\Sigma T_{k^*}^+ = (U\Sigma - \tilde{E}_k) T_k^+ W + \tilde{E}_{k^*} T_{k^*}^+ && \mid \cdot T_{k^*} \\
\Rightarrow\ & U\Sigma = (U\Sigma - \tilde{E}_k) T_k^+ W T_{k^*} + \tilde{E}_{k^*} \\
\Rightarrow\ & 0 = U\Sigma (T_k^+ W T_{k^*} - I) - \tilde{E}_k T_k^+ W T_{k^*} + \tilde{E}_{k^*}.
\end{aligned} \tag{5.6}$$

Weil $\mathrm{Im}(U\Sigma) \perp \mathrm{Im}(\tilde{E}_k)$ gilt, müssen der erste und die zwei weiteren Terme der rechten Seite in (5.6) einzeln betrachtet bereits null ergeben. Damit folgen die Behauptungen b) und c) durch

$$\begin{aligned}
0 = U\Sigma (T_k^+ S T_{k^*} - I) \ &\Rightarrow I = T_k^+ W T_{k^*} \ \Rightarrow\ W = T_k T_{k^*}^+ \\
&\Rightarrow 0 = \tilde{E}_{k^*} - \tilde{E}_k T_k^+ T_k T_{k^*}^+ T_{k^*} \Rightarrow\ 0 = \tilde{E}_{k^*} - \tilde{E}_k \\
&\Rightarrow \tilde{E}_{k^*} = \tilde{E}_k.
\end{aligned}$$

□

Durch Satz 6 kann nun eine Abschätzung des Fehlers der kinetischen Anpassung für beliebige $k \in \mathcal{K}(k^*)$ durchgeführt werden.

Folgerung 1. *Seien die Voraussetzungen wie in Satz 6. Mit $k \in \mathcal{K}(k^*)$ und dem Fehler $\delta = \|(C(k^*) - C^{\mathrm{dgl}}(k^*)) B^{-1}\|_F$ gilt die Ungleichung*

$$\|(C(k) - C^{\mathrm{dgl}}(k)) B^{-1}\|_F \leq \delta \|B(T_{k^*} T_k^+) B^{-1}\|_F \,.$$

Beweis:

$$\begin{aligned}
\|(C(k) - C^{\mathrm{dgl}}(k))B^{-1}\|_F &= \|(U\Sigma T_k^+ - C^{\mathrm{dgl}}(k^*)T_{k^*}T_k^+)B^{-1}\|_F \\
&= \|(U\Sigma T_{k^*}^+ T_{k^*} T_k^+ - C^{\mathrm{dgl}}(k^*)T_{k^*}T_k^+)B^{-1}\|_F \\
&= \|(U\Sigma T_{k^*}^+ - C^{\mathrm{dgl}}(k^*)(B^{-1}B)(T_{k^*}T_k^+)B^{-1}\|_F \\
&\leq \|(U\Sigma T_{k^*}^+ - C^{\mathrm{dgl}}(k^*)B^{-1}\|_F \, \|B(T_{k^*}T_k^+)B^{-1}\|_F \\
&= \delta \, \|B(T_{k^*}T_k^+)B^{-1}\|_F
\end{aligned}$$

□

Um die Ergebnisse der Folgerung praktisch nutzen zu können, wird eine obere Abschätzung von $\|B(T_{k^*}T_k^+)B^{-1}\|_F$ für alle $k \in \mathcal{K}(k^*)$ bestimmt.

Folgerung 2. *Seien die Voraussetzungen wie in Folgerung 1 und $y := (U\Sigma)^+(1,\ldots,1)^T$. Weiter seien die Einträge aller $C(k)$ mit $k \in \mathcal{K}(k^*)$ betragsmäßig kleiner als c_Σ und weiter c_Σ minimal. Dann gilt*

$$\|B(T_{k^*}T_k^+)B^{-1}\|_F \leq \sqrt{\sum_{i=1}^{s} \frac{1}{B_{i,i}^2}} \max_{\substack{x\in\mathbb{R}^s \\ |x|\leq c_\Sigma |y|}} \|BT_{k^*}x\|_2 \quad \forall k \in \mathcal{K}(k^*). \tag{5.7}$$

Beweis: Mit der vereinfachten Notation $X := T_k^+$ gelten

$$\begin{aligned}
\|B(T_{k^*}X)B^{-1}\|_F^2 &= \left\| BT_{k^*} \left(\frac{1}{B_{1,1}} X_{:,1}, \ \ldots, \ \frac{1}{B_{s,s}} X_{:,s} \right) \right\|_F^2 \\
&= \sum_{i=1}^{s} \frac{1}{B_{i,i}^2} \|BT_{k^*}X_{:,i}\|_2^2 \\
&\leq \left(\sum_{i=1}^{s} \frac{1}{B_{i,i}^2} \right) \max_{x\in\mathbb{R}^s} \|BT_{k^*}x\|_2^2 \ .
\end{aligned}$$

Nun wird noch die zusätzliche Einschränkung der zulässigen Vektoren x der Maximierung in (5.7) hergeleitet. Für relevante x gilt nach Voraussetzung

$$\begin{aligned}
& -c_\Sigma \, (1,\ldots,1)^T \leq U\Sigma x \leq c_\Sigma \, (1,\ldots,1)^T \\
\Rightarrow \quad & |(U\Sigma)^+ U\Sigma x| \leq c_\Sigma \, |(U\Sigma)^+(1,\ldots,1)^T| \\
\Rightarrow \quad & |x| \leq c_\Sigma \, |y| \ ,
\end{aligned}$$

wobei die Beträge komponentenweise anzuwenden sind.

□

Eine Lösung des Maximierungsproblems in Gleichung (5.7) kann direkt bestimmt werden. Wird die Zielfunktion

$$f(x) = \|BT_{k^*}x\|_2^2 = x^T \underbrace{(BT_{k^*})^T BT_{k^*}}_{A} x \tag{5.8}$$

zugrunde gelegt, so ist f wegen der positiven Definitheit von A konvex. Die Lösung der Maximierung von $f(x)$ ist dann in den Ecken des durch $|x| \leq c_\Sigma |y|$ definierten Suchgebiets (Hyperrechteck) zu finden. Eine Darstellung einer solchen Zielfunktion $f(x)$ ist in Abbildung 5.1 für ein rechteckiges Suchgebiet exemplarisch dargestellt.

Mit den Folgerungen 1 und 2 lässt sich nun der zu erwartende maximale Fehler $\delta_{\max}$, siehe (5.3), bei der Approximation von $\mathcal{K}_\varepsilon$ durch $\mathcal{K}(k^*)$ abschätzen.

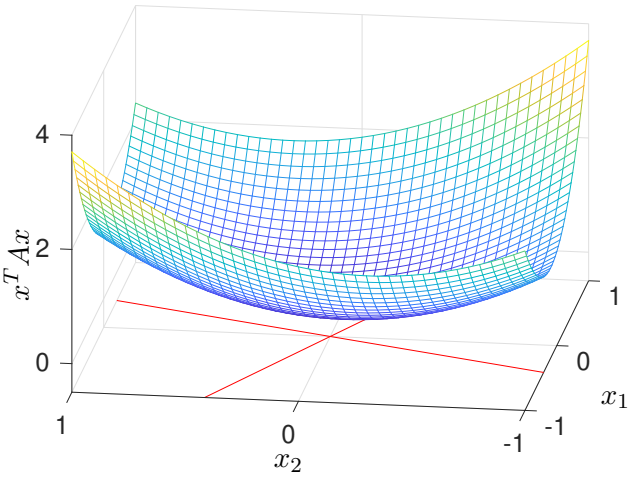


Abbildung 5.1.: Darstellung der Zielfunktion (5.8) im $\mathbb{R}^2$ für ein Modellproblem. Die Eigenräume der Matrix A werden durch die roten Linien repräsentiert.

5.2. Parameterlösungsmengen der Michaelis-Menten Kinetik

In diesem Abschnitt wird die Michaelis-Menten Kinetik

$$\mathcal{R} + \mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \xrightarrow{k_3} \mathcal{X} + \mathcal{P} \tag{5.9}$$

mit einem Reaktant $\mathcal{R}$, einem Katalysator $\mathcal{X}$, einem Katalysatorkomplex $\mathcal{Y}$ und einem Produkt $\mathcal{P}$ bezüglich der Existenz nichttrivialer Mengen D-approximativer Parameter $\mathcal{K}_\varepsilon$ untersucht. Das übergeordnete Ziel ist weiterhin die Bestimmung des Konzentrations- und Spektrenfaktors mit zugehöriger D-approximativer Parametrierung k^* von (5.9).

Die Michaelis-Menten Kinetik hat ihren Ursprung in der Biologie und dient zur Beschreibung enzymatischer Vorgänge, also Reaktionen bei der die Geschwindigkeit der Bildung des Produkts maßgeblich von der Konzentration eines Enzyms abhängig ist. Man spricht allgemein von katalysierten Reaktionen, wobei der Katalysator hierbei das Analogon zum Enzym ist.

Diese spezielle Kinetik wird betrachtet, weil sie ein interessantes Verhalten bezüglich der Existenz nichttrivialer Mengen $\mathcal{K}_\varepsilon$ in Abhängigkeit der gewählten Startkonzentrationen besitzt. Des Weiteren wird das kinetische Modell (5.9) zur Beschreibung von Reaktionssystemen unserer Kooperationspartner am LIKAT und in der Evonik Performance Materials GmbH eingesetzt, wodurch sich auch aus Anwendersicht ein besonderes Interesse ergibt. Der Abschnitt ist in drei Teile gegliedert: die Problembeschreibung, eine Herleitung der Menge $\mathcal{K}_\varepsilon$ und die Beschreibung eines möglichen Ansatzes zur eindeutigen Bestimmung von k^*.

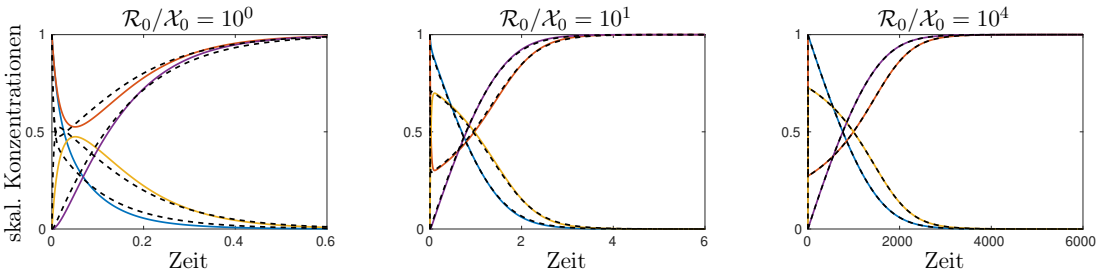


Abbildung 5.2.: Darstellungen von $C^{\mathrm{dgl}}_{\mathrm{skal}}(k^*)$ in bunt und $C^{\mathrm{dgl}}_{\mathrm{skal}}(k^{\mathrm{test}})$ in schwarz für unterschiedliche Verhältnisse $\mathcal{R}_0/\mathcal{X}_0$. In der rechten Grafik zum Fall $\mathcal{R}_0/\mathcal{X}_0 = 10^4$ ist kein Unterschied der beiden Faktoren mehr zu erkennen.

Problembeschreibung

Mit den Konzentrationsverläufen $\mathcal{R} := \mathcal{R}(t)$, $\mathcal{X} := \mathcal{X}(t)$, $\mathcal{Y} := \mathcal{Y}(t)$ und $\mathcal{P} := \mathcal{P}(t)$ sei durch

$$\begin{aligned} \dot{\mathcal{R}} &= -k_1\mathcal{R}\,\mathcal{X} + k_2\mathcal{Y} \\ \dot{\mathcal{X}} &= -k_1\mathcal{R}\,\mathcal{X} + k_2\mathcal{Y} + k_3\mathcal{Y} \\ \dot{\mathcal{Y}} &= \quad k_1\mathcal{R}\,\mathcal{X} - k_2\mathcal{Y} - k_3\mathcal{Y} \\ \dot{\mathcal{P}} &= \qquad\qquad\qquad\quad k_3\mathcal{Y} \end{aligned} \tag{5.10}$$

mit $(\mathcal{R}(0), \mathcal{X}(0), \mathcal{Y}(0), \mathcal{P}(0))^T = (\mathcal{R}_0, \mathcal{X}_0, 0, 0)^T$ das parameterabhängige Anfangswertproblem zu (5.9) definiert. Es wird im Folgenden ein Modellproblem betrachtet. Seien die im Normalfall unbekannte Kinetikparametrierung $k^* = (40, 5, 10)^T$ und der dadurch bestimmte Konzentrationsfaktor $C = C^{\mathrm{dgl}}(k^*)$ gegeben. Der nichtnegative Spektrenfaktor S ist dem Reaktionssystem aus Anhang A.2 nachempfunden, wobei er für die Betrachtungen dieses Abschnitts ohnehin keine entscheidende Rolle spielt.

Die Bestimmung von k^* aus einer entsprechenden Spektrenfolge D mittels der Ansätze aus Kapitel 3 ist für Startwerte mit $\mathcal{R}_0 \approx \mathcal{X}_0$ unproblematisch, denn es gilt, dass die modellierten Konzentrationsfaktoren $C^{\mathrm{dgl}}(k^*)$ und $C^{\mathrm{dgl}}(k')$ für $k^* \neq k'$ eindeutig unterschieden werden können. Der aus Anwendersicht relevantere Fall ist $\mathcal{R}_0 \gg \mathcal{X}_0$ mit $\mathcal{R}_0/\mathcal{X}_0 \geq 10^3$, für den aus $k^* \neq k'$ auch nahezu identische Konzentrationsfaktoren resultieren können. Diese Problematik ist in Abbildung 5.2 veranschaulicht. Die skalierten Konzentrationsfaktoren $C^{\mathrm{dgl}}_{\mathrm{skal}}(k) := C^{\mathrm{dgl}}(k) \cdot \mathrm{diag}(\mathcal{R}_0, \mathcal{X}_0, \mathcal{X}_0, \mathcal{R}_0)^{-1}, k \in \{k^*, k^{\mathrm{test}}\}$ sind für k^* (bunt) und eine zweite Parametrierung $k^{\mathrm{test}} = (200, 65, 10)^T$ (schwarz) zu $\mathcal{R}_0 = 1$ und $\mathcal{R}_0/\mathcal{X}_0 \in \{10^i : i \in \{0, 1, 4\}\}$ dargestellt. Die Differenz zwischen $C^{\mathrm{dgl}}_{\mathrm{skal}}(k^*)$ und $C^{\mathrm{dgl}}_{\mathrm{skal}}(k^{\mathrm{test}})$ ist für den Fall $\mathcal{R}_0/\mathcal{X}_0 = 1$ in der linken Grafik gut zu erkennen. Visuell sind Unterschiede zwischen den beiden Konzentrationsfaktoren für den Fall $\mathcal{R}_0/\mathcal{X}_0 = 10$ (mittig) bereits nur noch schwierig zu sehen und für $\mathcal{R}_0/\mathcal{X}_0 = 10^4$ (rechts) praktisch nicht mehr erkennbar. Die Beobachtung für den Fall $\mathcal{R}_0/\mathcal{X}_0 = 10^4$ wird ebenfalls durch den zugehörigen relativen Fehler

$$\mathrm{Err}_{\mathcal{X}_0=10^{-4}} = \left\| C^{\mathrm{dgl}}_{\mathrm{skal}}(k^*) - C^{\mathrm{dgl}}_{\mathrm{skal}}(k^{\mathrm{test}}) \right\|_F \Big/ \left\| C^{\mathrm{dgl}}_{\mathrm{skal}}(k^*) \right\|_F = 1.36 \cdot 10^{-5}$$

bestätigt. Die Bestimmung von k^* ist in diesem Fall mittels der numerischen Ansätze aus Kapitel 3 nicht realisierbar, weil ein Unterscheiden der Parametrierungen k^* und k^{test} anhand des Fehlers der Kinetikanpassung nicht sinnvoll möglich ist.

Dieses Problem wird zudem durch das Vorhandensein von Störungen weiter verstärkt. Hierzu ist in Abbildung 5.3 der Fehler $\mathrm{Err}_{\mathcal{X}_0}$ in Abhängigkeit von $\mathcal{X}_0$ über dem Inter-

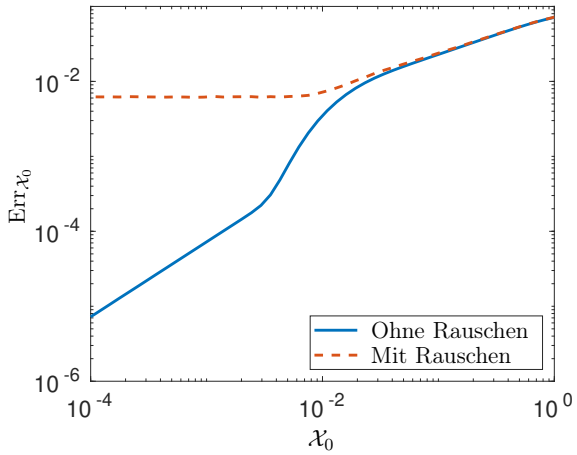


Abbildung 5.3.: Die Grafik zeigt den Fehler $\mathrm{Err}_{\mathcal{X}_0}$ in Abhängigkeit von $\mathcal{X}_0$ im Intervall $[10^{-4}, 1]$. Mit blau ist der störungsfreie und mit rot der durch homoskedastisches Rauschen von 0.5% gestörte Fall dargestellt. Das Rauschen wird vor der Bestimmung von $\mathrm{Err}_{\mathcal{X}_0}$ zu den Matrizen $C^{\mathrm{dgl}}_{\mathrm{skal}}(\cdot)$ hinzuaddiert. Ist $\mathcal{X}_0$ kleiner als etwa 10^{-2}, ist ein signifikanter Einfluss des Rauschens auf den Fehler $\mathrm{Err}_{\mathcal{X}_0}$ zu erkennen.

vall $[10^{-4}, 1]$ für den störungsfreien und einen gestörten Fall aufgetragen. Anhand des blauen Verlaufs zum ungestörten Fall ist zu erkennen, dass der relative Fehler $\mathrm{Err}_{\mathcal{X}_0}$ zwischen den skalierten Konzentrationsfaktoren mit kleiner werdendem $\mathcal{X}_0$ ebenfalls abnimmt. Dies ist konsistent zu den Aussagen des vorherigen Absatzes. Für den als roten Verlauf dargestellten Fehler des gestörten Falls, ist ab etwa $\mathcal{X}_0 = 10^{-2}$ kein weiteres Abfallen des Fehlers zu erkennen. Weil sich die beiden Fälle einzig durch die Störung mittels Rauschen unterscheiden, kann geschlussfolgert werden, dass die Differenz der ungestörten skalierten Konzentrationsfaktoren für $\mathcal{X}_0 < 10^{-2}$ signifikant kleiner ist als der Einfluss des Rauschens. Die wichtige Konsequenz ist, dass selbst für simulierte Daten unter dem Einfluss von Rauschen eine Unterscheidung von $C^{\mathrm{dgl}}_{\mathrm{skal}}(k^*)$ und $C^{\mathrm{dgl}}_{\mathrm{skal}}(k^{\mathrm{test}})$ nur für $\mathcal{X}_0 > 10^{-2}$ möglich ist.

Weil sich die Konzentrationsfaktoren für k^* und k^{test} offensichtlich für $\mathcal{X}_0 < 10^{-2}$ nur geringfügig unterscheiden, gilt für ein hinreichend großes (aber sinnvolles) ε, dass sowohl $k^* \in \mathcal{K}_\varepsilon$ als auch $k^{\mathrm{test}} \in \mathcal{K}_\varepsilon$ gelten. Es stellen sich die Fragen: Für welche $k \in \mathbb{R}^3_+$ neben k^{test} gilt $C^{\mathrm{dgl}}(k^*) \approx C^{\mathrm{dgl}}(k)$ und lässt sich daraus eine Approximation der Menge D-approximativer Parameter $\mathcal{K}_\varepsilon$ herleiten?

Nichttriviale Parameterlösungsmengen $\mathcal{K}_\varepsilon$

Es wird das Anfangswertproblem (5.10) mit $\mathcal{R}_0 = 1$ und $\mathcal{X}_0 = 10^{-4}$ zum Zeitgitter $T_G = [0, 3, \ldots, 6000]$ betrachtet. Für die Differenz der Konzentrationen der Komponente $\mathcal{R}$ bezüglich zwei beliebiger aufeinander folgender Stützstellen $t_i, t_{i+1} \in T_G$ gilt

$$|\mathcal{R}(t_i) - \mathcal{R}(t_{i+1})|/\mathcal{R}(t_i) \leq 0.008\,,$$

womit bezüglich des gewählten Gitters von einer langsamen Änderung der Konzentration $\mathcal{R}(t)$ ausgegangen werden kann. Es wird nun gezeigt, dass sich im Gegensatz dazu der Gleichgewichtszustand zwischen $\mathcal{X}$ und $\mathcal{Y}$ in Abhängigkeit von $\mathcal{R}$ schnell einstellt. Hierzu wird das, aus dem Anfangswertproblem (5.10) für eine fixierte Konzentration $\tilde{\mathcal{R}}$

hergeleitete, Differentialgleichungssystem betrachtet:

$$\begin{aligned} \dot{\mathcal{X}} &= -k_1\tilde{\mathcal{R}}\,\mathcal{X} + k_2\mathcal{Y} + k_3\mathcal{Y} \\ \dot{\mathcal{Y}} &= \quad k_1\tilde{\mathcal{R}}\,\mathcal{X} - k_2\mathcal{Y} - k_3\mathcal{Y} \\ \Leftrightarrow \quad \dot{\mathcal{X}} &= -k_1\tilde{\mathcal{R}}\mathcal{X} + (k_2+k_3)\mathcal{Y} \\ \dot{\mathcal{Y}} &= \quad k_1\tilde{\mathcal{R}}\mathcal{X} - (k_2+k_3)\mathcal{Y} \end{aligned}. \tag{5.11}$$

Weiter seien hierzu der stationäre Punkt $(\mathcal{X}_G, \mathcal{Y}_G)$ von (5.11) und die Startkonzentrationen $(\mathcal{X}_\beta, \mathcal{Y}_\beta) = (\beta\mathcal{X}_G, (1-\beta)\mathcal{X}_G + \beta\mathcal{Y}_G)$ mit $\beta \in [0, (\mathcal{X}_G + \mathcal{Y}_G)/\mathcal{X}_G]$ gegeben. Mit dem gewählten β gilt $0 \leq \mathcal{X}_\beta, \mathcal{Y}_\beta \leq \mathcal{X}_0$, womit $(\mathcal{X}_\beta, \mathcal{Y}_\beta)$ die Werte aller relevanten Fälle annehmen kann. Die Lösung $(\mathcal{X}(t), \mathcal{Y}(t))$ des Anfangswertproblems aus (5.10) mit Startkonzentrationen $(\mathcal{X}_\beta, \mathcal{Y}_\beta)$ ist durch die Gleichung (4.21) gegeben. Um abschätzen zu können, wie schnell der Gleichgewichtszustand erreicht wird, kann der Punkt

$$\left(\mathcal{X}_\lambda, \mathcal{Y}_\lambda\right) := (1-\lambda)\left(\mathcal{X}_\beta, \mathcal{Y}_\beta\right) + \lambda\left(\mathcal{X}_G, \mathcal{Y}_G\right)$$

mit $\lambda \in [0,1]$ betrachtet werden. Auflösen von $\mathcal{X}_\lambda = \mathcal{X}(t)$ nach t ergibt

$$t(\lambda) = \frac{\ln((1-\lambda)^{-1})}{\tilde{\mathcal{R}}k_1 + k_2 + k_3},$$

womit die Angabe der Zeit bis zu einem durch λ beschriebenen Fortschritt der modellierten Konzentrationen von $(\mathcal{X}_\beta, \mathcal{Y}_\beta)$ nach $(\mathcal{X}_G, \mathcal{Y}_G)$ möglich ist. Wird die Parametrierung k^* und $\lambda = 0.99$ betrachtet, ist $t(0.99) \approx 0.084$ und somit deutlich kleiner als die Schrittweite von T_G, welche 3 beträgt.

Für den stationären Punkt $(\mathcal{X}_G, \mathcal{Y}_G)$ gilt weiter

$$\frac{\mathcal{X}_G}{\mathcal{Y}_G} = \frac{k_2+k_3}{\tilde{\mathcal{R}}k_1} = \frac{1}{\tilde{\mathcal{R}}}\underbrace{\frac{k_2+k_3}{k_1}}_{K_m} \tag{5.12}$$

mit der sogenannten Michaeliskonstante K_m. Die Vorbetrachtungen des letzten Absatzes rechtfertigen außerdem die Annahme, dass (5.12) approximativ für fast alle Zeitpunkte t bezüglich des ursprünglichen Modells (5.10), das heißt

$$\mathcal{R}(t)\mathcal{X}(t) = K_m\mathcal{Y}(t), \tag{5.13}$$

gilt. Ausgenommen ist $t = 0$, weil ein Einstellen des Gleichgewichtszustands zwischen $\mathcal{X}$ und $\mathcal{Y}$ noch nicht möglich ist.

Im vorherigen Abschnitt wurde beobachtet, dass die Matrizen $C^{\text{dgl}}_{\text{skal}}(k)$ zu verschiedenen Parametrierungen k für das Anfangswertproblem dieses Abschnitts nahezu übereinstimmen. Nun erfüllen die zugrunde liegenden Konzentrationsverläufe aber die Gleichung (5.13), woraus folgt, dass die besagten Parametrierungen dieselbe Michaeliskonstante K_m haben müssen. Es stellt sich die Frage, ob sich die aus k^* resultierende Michaeliskonstante K_m auch aus anderen Parametrierungen $k' = (k'_1, k'_2, k'_3)$ ergibt. Gelten $k'_3 = k_3$ und der lineare Zusammenhang $k'_2 = k'_2(k'_1) := K_m k'_1 - k'_3$, folgen die Gleichungen

$$\frac{k'_2 + k'_3}{k'_1} = \frac{K_m k'_1 - k_3 + k'_3}{k'_1} = \frac{K_m k'_1}{k'_1} = K_m = \frac{k_2+k_3}{k_1}.$$

Dies kann für das anfängliche Beispiel mit $k^* = (40,5,10)^T, k^{\text{test}} = (200,65,10)^T$ und daraus resultierend $K_m = (5+10)/40 = (65+10)/200 = 3/8$ verifiziert werden. Es ist

festzustellen, dass genau die Parametrierungen auf fast identische Konzentrationsfaktoren führen, für die der eben beschriebene lineare Zusammenhang gilt. Eine Approximation der Menge $\mathcal{K}_\varepsilon$ mit $K_m = (k_2^* + k_3^*)/k_1^*$ lautet also

$$\tilde{\mathcal{K}}_\varepsilon = \{k \in \mathbb{R}_+^3 : k_3 = k_3^* \wedge k_2 = K_m k_1 - k_3\}$$

und ist links in Abbildung 5.4 exemplarisch dargestellt. In der Grafik ist die k_1'-Achse auf das Intervall $[0, 1800]$ begrenzt. Tatsächlich kann der gezeigte Strahl beliebig in positiver Richtung fortgesetzt werden. Die mittlere Grafik zeigt die Spalten der Matrizen $C_{\text{skal}}^{\text{dgl}}(k)$ zu den in der linken Grafik mit blauen Kreuzen markierten Parametrierungen k. Visuell ist kein Unterschied zwischen den Matrizen zu erkennen. Die maximale Differenz zwischen zwei beliebigen der dargestellten Matrizen beträgt

$$\frac{\|C_{\text{skal}}^{\text{dgl}}(k^*) - C_{\text{skal}}^{\text{dgl}}([1800, 665, 10])\|_F}{\|C_{\text{skal}}^{\text{dgl}}(k^*)\|_F} \leq 1.131 \cdot 10^{-5} \ .$$

Unterschiede zwischen den numerischen Lösungen der Anfangswertprobleme sind erst bei starker Erhöhung der Auflösung und durch Betrachten des Zeitintervalls $[0, 0.1]$ möglich, wie in der rechten Grafik zu erkennen ist. Könnte diese Differenz zwischen zwei Faktoren gemessen werden, wäre auch eine Unterscheidung der zugehörigen Geschwindigkeitsparameter möglich. Die Einstellung des Gleichgewichts (5.12) zu Beginn der Reaktion lässt sich aber typischerweise nicht gut spektroskopisch vermessen, weil gerade hier diverse Störungen durch Mischungseffekte zwischen den Komponenten auftreten. Eine Unterscheidung der Parametrierungen lässt sich auf diesem Weg häufig nicht erreichen.

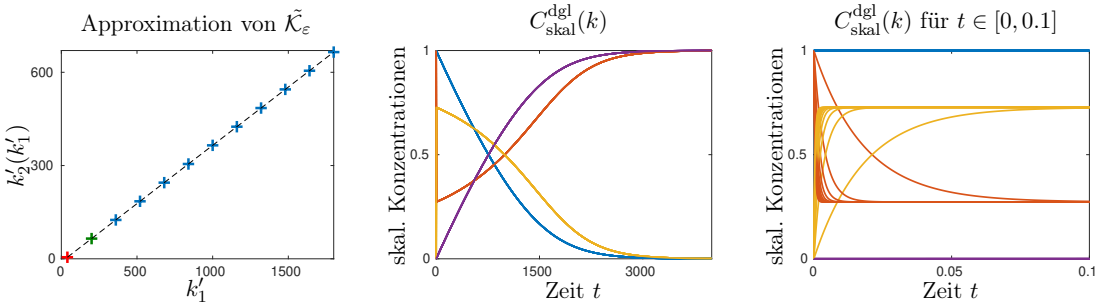


Abbildung 5.4.: Links ist ein Ausschnitt der Menge $\tilde{\mathcal{K}}_\varepsilon$ dargestellt. Der Punkt k^* ist als rotes und k^{test} als grünes Kreuz markiert. Weiter sind in der mittleren Grafik die Matrizen $C_{\text{skal}}^{\text{dgl}}(k)$ für alle in der linken Grafik als Kreuz dargestellten Punkte mit $k_3' = 10$ gezeigt. In der Mitte wird ein Zeitgitter, welches einem typischen Spektrometer nachempfunden ist, genutzt, wohingegen rechts die Einschränkung auf das Zeitintervall $[0, 0.1]$ mit einer stark erhöhten Auflösung verwendet wird.

Ein Multiset-Lösungsansatz

Kann durch Veränderung der Reaktionsbedingungen wie Druck oder Temperatur Einfluss auf die Geschwindigkeitsparameter genommen werden, ist eine eindeutige Bestimmung von k^* dennoch häufig möglich. In dieser Fallstudie wird davon ausgegangen, dass eine Skalierung von k_3 durch einen bekannten und steuerbaren Faktor $\zeta > 0$ möglich ist. Das Differentialgleichungssystem hat dann die Form:

$$\begin{aligned}
\dot{\mathcal{R}} &= -k_1 \mathcal{R}\,\mathcal{X} + k_2 \mathcal{Y} \\
\dot{\mathcal{X}} &= -k_1 \mathcal{R}\,\mathcal{X} + k_2 \mathcal{Y} + \zeta k_3 \mathcal{Y} \\
\dot{\mathcal{Y}} &= \ \ k_1 \mathcal{R}\,\mathcal{X} - k_2 \mathcal{Y} - \zeta k_3 \mathcal{Y} \\
\dot{\mathcal{P}} &= \qquad\qquad\qquad\quad \zeta k_3 \mathcal{Y}
\end{aligned}$$

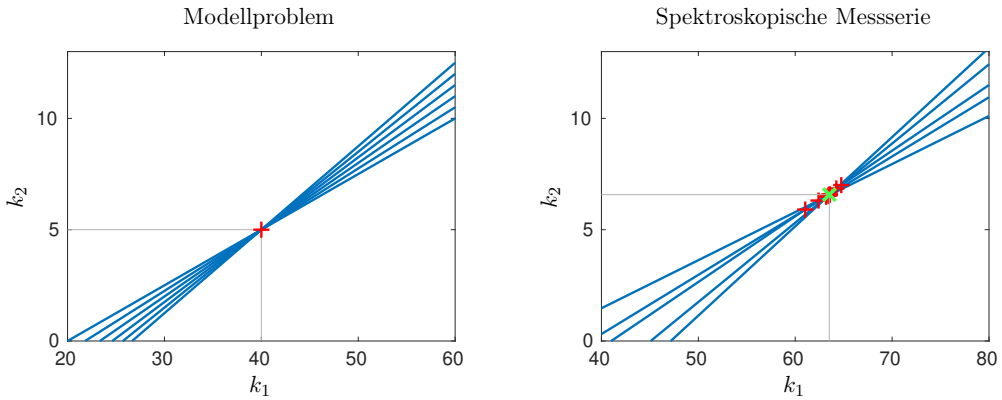


Abbildung 5.5.: Links sind die Approximationen der Mengen $\mathcal{K}_\varepsilon$ für die Fallstudie des Abschnitts 5.2 als blaue Geraden dargestellt. Rechts ist der gleiche Zusammenhang für einen spektroskopischen Datensatz zu sehen. In beiden Grafiken markieren die roten Kreuze die Schnittpunkte der Geraden miteinander. Für den spektroskopischen Datensatz kann erwartungsgemäß kein eindeutiger Schnittpunkt festgestellt werden. Stattdessen verteilt sich die Menge an Schnittpunkten in einem kleinen Bereich. Der im Sinne kleinster Fehlerquadrate berechnete Geschwindigkeitsparameter k^* ist als grünes Kreuz eingezeichnet.

Die Berechnung einer Faktorisierung mit simultaner Anpassung einer Kinetik ermöglicht lediglich die Bestimmung von $\bar{k}_3 := \zeta k_3$, wobei k_3 wegen der bekannten Werte für ζ ebenfalls zugänglich ist. Es zeigt sich sogar, dass die Kenntnis von $\bar{k}_3$ ausreichend ist. Die von ζ abhängige Michaeliskonstante $\bar{K}_m := K_m(\zeta) = (k_2 + \zeta k_3)/k_1 = (k_2 + \bar{k}_3)/k_1$ bestimmt den Anstieg der Geraden $k_2' = k_2'(k_1') := \bar{K}_m k_1' - \bar{k}_3$ und somit die Approximation der Menge $\mathcal{K}_\varepsilon$ (vergleiche Abbildung 5.4) für das jeweilige ζ. Stehen mindestens zwei Datensätze zu verschiedenen Werten ζ zur Verfügung, haben die entsprechenden Geraden unterschiedliche Anstiege sowie Absolutterme, womit sich ein eindeutig bestimmter Schnittpunkt ergibt. Seien hierzu zwei dieser Geraden zu ζ_1 und ζ_2 durch die Gleichungen $k_2' = K_m(\zeta_1)k_1' - \zeta_1 k_3$ und $k_2' = K_m(\zeta_2)k_1' - \zeta_2 k_3$ gegeben. Ihr Schnittpunkt ist unabhängig von ζ_1 und ζ_2, lautet (k_1, k_2) und entspricht den ersten zwei Komponenten des gesuchten Geschwindigkeitsparameters k^*.

Es folgt nun die Anwendung des beschriebenen Vorgehens für zwei Beispiele. Zuerst wird ein Modellproblem betrachtet, welches auf der Fallstudie dieses Abschnitts mit $\zeta \in \{0.5, 0.6, \ldots, 1\}$ basiert. Die resultierenden Approximationen der Mengen $\mathcal{K}_\varepsilon$ sind als Geraden in der linken Grafik der Abbildung 5.5 dargestellt. Die Schnittpunkte der Geraden untereinander entsprechen dem Punkt (k_1^*, k_2^*), welcher zuvor nicht eindeutig bestimmbar war.

Als nächstes wird eine Folge von spektroskopischen Datensätzen betrachtet, die in [68] beschrieben ist. Die einzelnen Datensätze unterscheiden sich durch Variation des verwendeten Wasserstoff-Partialdrucks (hier ζ). Für einen jeden Datensatz wurde eine zulässige Zerlegung unter Berücksichtigung der Kinetik (5.10) berechnet. Die Vorgehensweise erfolgt dabei analog zu Anhang A.2. Die für diese Analyse relevanten Informationen über die extrahierten Geschwindigkeitsparameter sind auf Seite 134 der referenzierten Arbeit zu finden. Nach Rücksprache mit Dr. C. Kubis wurden die Geschwindigkeitsparameter zu niedrigen Wasserstoffkonzentrationen aufgrund geringerer Güte der Ergebnisse vernachlässigt. Die Wasserstoffkonzentrationen entsprechen hierbei den skalaren Faktoren $\zeta_1, \ldots, \zeta_5$. Die daraus bestimmten Approximationen der Mengen $\mathcal{K}_\varepsilon$ sind in der rechten Grafik der Abbildung 5.5 zu sehen. Die Menge der Schnittpunkte der Geraden konzen-

triert sich auf einen kleinen Bereich. Durch Lösen des Ausgleichssystems

$$\begin{pmatrix} K_m(\zeta_1) & -1 \\ \vdots & \vdots \\ K_m(\zeta_5) & -1 \end{pmatrix} \begin{pmatrix} k_1^* \\ k_2^* \end{pmatrix} = \begin{pmatrix} \zeta_1 k_3 \\ \vdots \\ \zeta_5 k_3 \end{pmatrix}$$

lässt sich mit $(k_1^*, k_2^*)^T = (63.52\text{min}^{-1}, 6.58\text{min}^{-1})^T$ eine Approximation berechnen. Sie ist als grünes Kreuz in der Grafik dargestellt. Dies zeigt, dass dieser Ansatz auch für spektroskopische störungsbehaftete Messserien einsetzbar ist.

5.3. L-Kurven zur Wahl von Gewichtungen

Die Gewichtung der in den Kapiteln 2 und 3 genutzten Zielfunktionen hat für störungsbehaftete Daten teils erhebliche Auswirkung auf die berechneten Lösungen der entsprechenden Matrixfaktorisierungsaufgaben, wie beispielsweise für den spektroskopischen Datensatz 1 aus Kapitel 3 gezeigt wurde. Um den Zusammenhang der Gewichte einer solchen Zielfunktion zu untersuchen, können sogenannte L-Kurven [15, 16, 56, 57] eingesetzt werden. Dabei werden die Residuennormen der Zielfunktionsterme für variierte Gewichtungsverhältnisse grafisch gegeneinander aufgetragen. Daraus können Aussagen über die Sensitivität einer Matrixfaktorisierungen bezüglich der Wahl der Gewichtung getroffen werden. Auch die Approximation einer optimalen Gewichtung im Sinne eines ausgeglichenen Beitrags aller Zielfunktionsterme ist möglich. Diese Art der Analyse geht auf das Lösen von linearen Gleichungssystemen zu schlecht konditionierten Koeffizientenmatrizen unter Verwendung der sogenannten Tikhonov-Regularisierung [94, 124] zurück.

Es wird exemplarisch die Zielfunktion (3.4) mit den Gewichten γ_3 und δ_1 analysiert, wobei nur die Nebenbedingungen der Nichtnegativität des Faktors S und die Konsistenz des Faktors C mit einem kinetischen Modells betrachtet werden. Die Nichtnegativität des Faktors C und die Regularität der Matrix T folgen direkt aus einer hinreichend guten Kinetikanpassung und werden vernachlässigt. Die verwendete Zielfunktion lautet dann

$$F_L(k) = \gamma_3 \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{s}\left(\min\left(\frac{S_{ij}}{\max_l(S_{lj})}, 0\right)\right)^2}_{\text{err}_S} + \delta_1 \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{s}\left(\frac{C_{i,j} - (C^{\text{dgl}}(k))_{i,j}}{\max_l(C_{l,j})}\right)^2}_{\text{err}_{\text{kin}}}, \quad (5.14)$$

wobei T, C und S wie für (3.6) berechnet werden. Weiter wird $\gamma_3 = \mu$ und $\delta_1 = 1 - \mu$ mit $\mu \in [0, 1]$ gewählt, womit die Wahl der Gewichtung der Zielfunktion auf den Parameter μ reduziert ist.

Zur Bestimmung einer L-Kurve wird der Parameter μ entlang eines Gitters variiert, jeweils eine Minimierung der entsprechenden Zielfunktion (5.14) durchgeführt und abschließend die so erhaltenen Residuennormen err_S und err_{kin} gegeneinander aufgetragen. In Abbildung 5.6 sind verschiedene L-Kurven zum spektroskopischen Datensatz 1 mit der Matrix $D \in \mathbb{R}^{735\times325}$ gezeigt. In der mittleren Zeile sind L-Kurven bezüglich der unveränderten Matrix D zu sehen. Die in den oberen Grafiken gezeigten L-Kurven basieren auf einer spaltenweise ausgedünnten Matrix D, welche nur jeden 5-ten Wellenzahlkanal enthält. Für die Grafiken der unteren Zeile wird analog bezüglich der Zeitachse verfahren und nur jede 15-te Zeile von D betrachtet. Im Gegensatz zur linken Spalte fließen für die rechts dargestellten L-Kurven die Dimensionen der Matrizen C und S in die Wahl der Gewichtungen mit $\gamma_3 = \mu\frac{m+n}{n}$ und $\delta_1 = (1-\mu)\frac{m+n}{m}$ ein.

Die folgenden Ergebnisse können der Abbildung 5.6 entnommen werden:

1. Es werden zuerst die linken drei Grafiken betrachtet bei denen die Gewichtung in (5.14) unabhängig von den Dimensionen m und n erfolgt. Die Verteilung der eingetragenen Hilfspunkte für $\mu \in \{0.6, 0.8, 0.9\}$ entlang der L-Kurven hängt wie erwartet vom Verhältnis m/n ab. Das bedeutet ein Ausdünnen von D kann bei gleichbleibender Gewichtung in unterschiedlichen Ergebnissen resultieren. Weiter weisen die drei Kurven eine lokal begrenzte starke Krümmung auf. Kleine Änderungen der Gewichte in diesem Bereich haben eine große Auswirkung auf die berechnete Matrixfaktorisierung. Im Gegensatz dazu ist abseits dieser starken Krümmung mit einer kleineren Auswirkung zu rechnen. Ohne Kenntnis der L-Kurven kann die Wahl einer Gewichtung somit schwierig sein. Für eine automatisierte Bestimmung von γ_3 und δ_1 ist dieser Umstand von Vorteil, weil beispielsweise eine Auswertung von L-Kurven bezüglich ihres Punktes größter Krümmung einfacher ist [58].
2. Für die Berechnung der L-Kurven in den rechten Grafiken werden die Dimensionen m und n berücksichtigt. Dies resultiert unter anderem in einer ähnlichen Verteilung der Hilfspunkte entlang der Kurven, woraus folgt, dass der Einfluss des Ausdünnens von D auf die Ergebnisse geringer ist. Weiter ist eine gleichmäßige Krümmung der Kurven festzustellen, wodurch eine manuelle Justierung der Gewichte einfacher ist, aber eine automatisierte Bestimmung erschwert wird.

Die hier exemplarisch am spektroskopischen Datensatz 1 durchgeführte Analyse zur optimalen Gewichtungswahl anhand von L-Kurven lässt sich in ganz analoger Weise auch auf andere Spektrenfolgen oder auch Zielfunktionen übertragen und bietet wertvolle Informationen für die numerische Lösung der Matrixfaktorisierungsprobleme 1 bis 4.

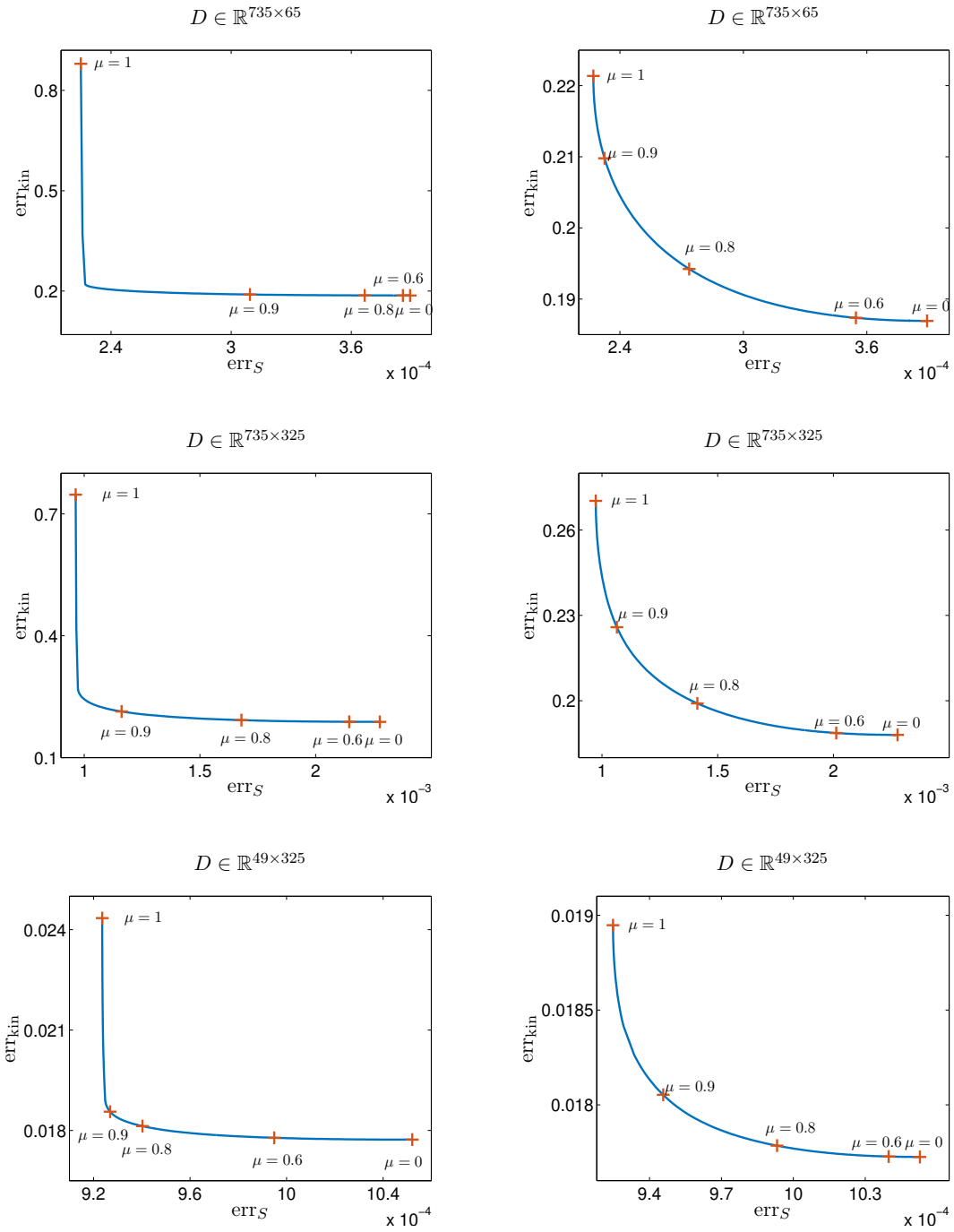


Abbildung 5.6.: Gezeigt sind L-Kurven zum spektroskopischen Datensatz 1, wobei dieser für die obere Zeile bezüglich der Wellenzahlen und für die untere Zeile bezüglich der Zeitachse ausgedünnt wurde. In der mittleren Zeile sind die L-Kurven bezüglich der unveränderten Matrix D dargestellt. Im Gegensatz zur linken wurden in der rechten Spalte zusätzlich die Dimensionen der Matrizen C und S bei der Wahl der Gewichte γ_3 und δ_1 beachtet. Die Details sind Abschnitt 5.3 zu entnehmen. Auf den Achsen sind jeweils die ungewichteten Residuennormen err_S und err_{kin}, siehe Gleichung (5.14), abgetragen. Ohne das Einbeziehen der Dimensionen der Matrix D in die Zielfunktion ist die Verteilung der Residuennormen err_S und err_{kin} erkennbar sensitiver bezüglich der durch μ definierten Gewichtung.

6. Numerische Approximation von Parameterlösungsmengen

Die Mengen $\mathcal{K}$, $\mathcal{K}^+$, $\mathcal{K}^+_{\varepsilon,\theta}$ und $\mathcal{K}_\varepsilon$ aus den Kapiteln 4 und 5 können als Niveaumengen

$$\mathcal{N} := \mathcal{N}_f(\alpha) = \{k \in \mathcal{G} : f(k) = \alpha\} \tag{6.1}$$

mit geeigneten Funktionen

$$f : \mathcal{G} \to \mathbb{R}_{\geq 0},\ \mathcal{G} \subseteq \mathbb{R}^q, \tag{6.2}$$

zu einem beschränkten Suchgebiet $\mathcal{G}$ und $\alpha = 0$ dargestellt werden. Exemplarisch seien die Funktionen $F_{\text{eig}}(k)$ in (4.14) für $\mathcal{K}$ und $F_\varepsilon(k)$ in (5.1) für $\mathcal{K}_\varepsilon$ genannt. Das Ziel dieses Kapitels ist die Herleitung geeigneter numerischer Methoden zur Approximation solcher Niveaumengen. Hierbei soll möglichst kein Vorwissen über die Dimension oder die genaue Struktur von $\mathcal{N}$ nötig sein. Es darf lediglich angenommen werden, dass ein $k^0 \in \mathcal{N}$ bekannt ist.

In Abschnitt 6.1 werden zwei in der Literatur beschriebene Methoden vorgestellt, die auf der Auswertung der Funktion f auf einem vorgegebenen, hinreichend feinen Gitter basieren. Typischerweise erreichen solche Ansätze nur unter großem Rechenaufwand eine akzeptable Approximationsgüte. Abschnitt 6.2 umfasst den Hauptteil des Kapitels. Es wird ein Algorithmus beschrieben, der iterativ eine Menge von Würfeln generiert, die $\mathcal{N}$ einschließen. Hierbei wird f mittels numerischer Verfahren minimiert, wobei lokal stark begrenzte Suchgebiete genutzt werden. Hierdurch sind typischerweise nur wenige Iterationen der Minimierungsverfahren nötig. So kann neben einer hohen Genauigkeit auch ein, im Verhältnis zu den anderen Methoden, geringer Rechenaufwand erreicht werden.

6.1. Gitterbasierte Methoden

Die nun vorgestellten Algorithmen basieren auf der Auswertung der Funktion f der Niveaumenge $\mathcal{N}$ auf einem Gitter. Der in Abschnitt 6.1.1 vorgestellte Grid Search Algorithmus nutzt hierbei ein festgelegtes Gitter und der Box Algorithmus aus Abschnitt 6.1.2 eine zusätzliche iterative Verfeinerungsstrategie.

6.1.1. Grid Search

Der Grid Search Algorithmus [47,78,98,127] in seiner Grundform ermöglicht die Analyse des Verhaltens einer Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ in einem beschränkten Gebiet. Die Grundidee besteht in der Festlegung von unteren und oberen Grenzen $k_u \in \mathbb{R}^2$ und $k_o \in \mathbb{R}^2$ für den Parameter k von $f(k)$. Hierdurch wird die Menge $\{k \in \mathbb{R}^2 : k_u \leq k \leq k_o\}$ definiert, welche möglichst die Menge $\mathcal{G}$ beinhaltet und durch ein Gitter G_d diskretisiert wird. Für

jeden Gitterpunkt $k_G \in G_d$ wird der Funktionswert $f(k_G)$ bestimmt. Diese Funktionswerte können nun beispielsweise als Oberflächenplot bezüglich des Gitters dargestellt und analysiert werden.

Eine Verallgemeinerung des Algorithmus sieht das Betrachten von Funktionen wie in (6.2) vor. Eine grafische Auswertung ist dann typischerweise schwierig beziehungsweise nicht mehr möglich. Weiter ist anzumerken, dass auf einem Gitter mit n_G Stützstellen je Dimension die Anzahl der benötigten Funktionsauswertungen n_G^q beträgt. Eine sinnvolle Anwendung des Algorithmus mit hinreichend feinem Gitter ist normalerweise nur möglich, wenn wenige Parameter und eine schnell auszuwertende Funktion $f(k)$ vorliegen.

Eine Schwierigkeit bei der Anwendung von Grid Search zur Bestimmung von $\mathcal{K}$ und $\mathcal{K}^+$ liegt darin, dass viele der in Abschnitt 4.1.2 gezeigten Strukturen, wie Mengen von Punkten oder Geraden, nur erfasst werden können, wenn diese auf oder sehr nahe an den Gitterpunkten der gewählten Diskretisierung liegen. Bei geringen Abweichungen werden die Strukturen entweder nicht erfasst oder es müssen teils große Fehlertoleranzen für $f(k) = 0$ einbezogen werden. In Abschnitt 6.3.1 erfolgt hierzu eine Veranschaulichung an einem Beispiel.

Zur Verbesserung der Genauigkeit können die Gitterpunkte zusätzlich als Startiterierten einer Minimierung von f genutzt werden [33]. Ein entsprechender Algorithmus wird in Anhang A.1 vorgestellt.

6.1.2. Box-Algorithmus

Der Box-Algorithmus wird im Programmpaket GAIO (Global analysis of invariant objects) genutzt [28], welches der Untersuchung des transienten Verhaltens und der geometrischen Eigenschaften von dynamischen Systemen dient. Die Idee besteht darin eine zu untersuchende Menge mit Quadern zu überdecken und die relevanten Eigenschaften der Menge innerhalb dieser Quader zu bestimmen. Sind hierzu iterative Verfahren nötig, resultiert eine solche Reduktion auf Teilprobleme häufig in einer geringen Anzahl an nötigen Iterationsschritten. Weiter können adaptive Verfeinerungsstrategien genutzt werden, um eine Erhöhung der Genauigkeit zu erreichen.

Die Idee kann wie folgt auf die Berechnung von $\mathcal{K}$, $\mathcal{K}^+$, $\mathcal{K}_\varepsilon$ oder $\mathcal{K}^+_{\varepsilon,\theta}$ übertragen werden. Unter Voraussetzung einer beschränkten Menge $\mathcal{G}$ wird diese durch gleichgroße, disjunkte Quader (oder allgemeiner Hyperrechtecke) überdeckt. Für jeden Quader wird geprüft, ob ein Punkt k innerhalb des Quaders existiert, sodass $f(k) = 0$ gilt. Dieser Schritt erfolgt durch Auswerten festgelegter Punkte innerhalb des jeweiligen Quaders [28]. Alle Quader, die diese Bedingung erfüllen, werden halbiert und wiederum geprüft. Der Box-Algorithmus kann somit als adaptive Variante des Grid Search Algorithmus aus Abschnitt 6.1.1 angesehen werden. Die vorgeschlagene Prüfung eines Quaders durch eine jeweils vorgegebene Menge an Punkten, führt zu den gleichen Schwierigkeiten wie bei dem Grid Search Algorithmus.

6.2. Würfeleinschließungsalgorithmus

Der nun vorgestellte Algorithmus vereint die Ideen, der Algorithmen aus dem vorherigen Abschnitt 6.1. Darüber hinaus ist er inspiriert durch den *triangle enclosure algorithm* [47] zur Approximation der Ränder der Segmente der Menge zulässiger Lösungen mittels einer iterativ bestimmten Folge von gleichseitigen Dreiecken. Im Gegensatz dazu wird im

Würfeleinschließungsalgorithmus das Ziel der Konstruktion einer Überdeckung der Niveaumenge $\mathcal{N}$ durch q-dimensionale Würfel einer vorgegebenen Seitenlänge ω verfolgt. Hierzu wird mit einem Würfel W_0, der ein bekanntes Element $k^0 \in \mathcal{N}$ der Niveaumenge enthält, begonnen. Sukzessive werden der initialen Menge $\{W_0\}$ weitere Würfel hinzugefügt bis $\mathcal{N}$ vollständig durch die Vereinigung dieser Würfel überdeckt ist. Für jeden hinzuzufügenden Würfel wird geprüft, ob er ein Element der Menge $\mathcal{N}$ enthält, indem eine Minimierung von f, siehe (6.2), durchgeführt wird. Für jede Optimierung wird das Minimum innerhalb des jeweiligen Würfels bestimmt, wodurch sich verschiedene Vorteile ergeben:

- Die räumliche Trennung der durchzuführenden Optimierungen ermöglicht die Parallelisierung des Algorithmus und reduziert somit die benötigte Zeit auf modernen Desktop-Computern um einen Faktor von 2 bis 16 je nach Anzahl der verwendeten CPU-Kerne.
- Die geometrische Struktur von $\mathcal{N}$ kann durch die Nachbarschaftsbeziehungen der Würfel analysiert werden.
- Jeder Würfel kann durch die finale Iterierte der jeweiligen Optimierung repräsentiert werden. Bereits durch wenige Optimierungsschritte kann so eine hohe Genauigkeit der Approximation von $\mathcal{N}$ erreicht werden.

In Algorithmus 1 wird die Berechnung einer Einschließung von $\mathcal{N}$ durch Würfel zu einer festen Seitenlänge ω beschrieben. In den folgenden Abschnitten werden die verwendete Datenstruktur und die entscheidenden Subroutinen *Nachbarn(W)*, *Schnelltest(W)* sowie *Zulässigkeitstest(W, ε)* detailliert erläutert. Abbildung 6.1 veranschaulicht eine Iteration von Algorithmus 1. Eine Implementierung in Matlab unter Verwendung der Optimization-Toolbox liegt dieser Arbeit bei.

Algorithmus 1 Würfeleinschließung

Input: bekanntes Element der Niveaumenge $k^* \in \mathcal{N}$, Würfelkantenlänge $\omega \in \mathbb{R}$, Fehlertoleranz $\varepsilon \geq 0$
Output: Würfeleinschließung $\mathcal{W}$ von $\mathcal{N}$

Konstruiere initialen Würfel W_0 mit Mittelpunkt k^* und Kantenlänge ω
$\mathcal{W} = \{W_0\}$
$\mathcal{C} = \text{Nachbarn}(\mathcal{W})$
$\mathcal{A} = \mathcal{W} \cup \mathcal{C}$
while $\mathcal{C} \neq \emptyset$ **do**
  $\mathcal{C}^+ = \emptyset$
  for all $W \in \mathcal{C}$ **do**
    if Schnelltest(W) **then**
      if Zulässigkeitstest(W, ε) **then**
        $\mathcal{W} = \mathcal{W} \cup \{W\}$
        $\mathcal{C}^+ = \mathcal{C}^+ \cup \{W\}$
      end if
    end if
  end for
  $\mathcal{C} = \text{Nachbarn}(\mathcal{C}^+) \setminus \mathcal{A}$
  $\mathcal{A} = \mathcal{A} \cup \mathcal{C}$
end while

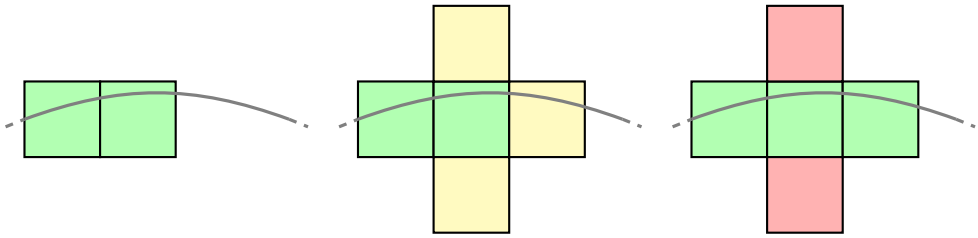


Abbildung 6.1.: Veranschaulichung der Bestimmung und Prüfung von zu testenden Würfeln (Kandidaten). Alle Würfel, welche benachbart zu Würfeln der Menge $\mathcal{W}$ (grün) liegen und selbst noch nicht Kandidat waren, bilden die Menge $\mathcal{C}$ (gelb). Sie werden anschließend, wie in Abbildung 6.2 gezeigt, geprüft und entweder akzeptiert (hinzugefügter grüner Würfel) oder verworfen (rot).

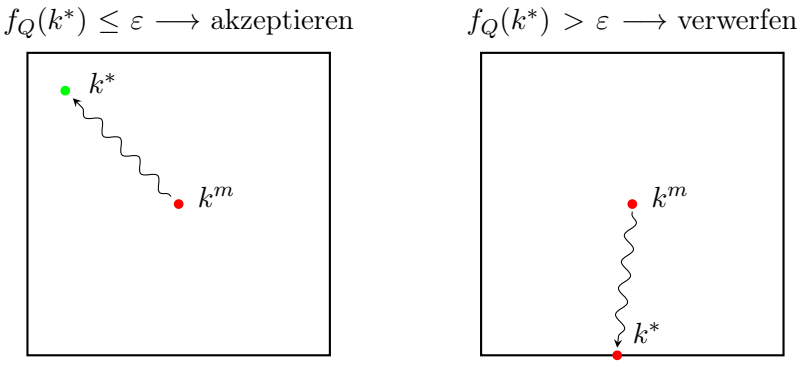


Abbildung 6.2.: Veranschaulichung des Zulässigkeitstests eines Würfels für $q = 2$. Beginnend beim Mittelpunkt k^m des Würfels wird eine Minimierung der Funktion $f_Q(k)$ (siehe (6.3)) durchgeführt, welche den optimalen Wert k^* innerhalb des Würfels bestimmen soll. Gilt für den entsprechende Funktionswert $f_Q(k^*) \leq \varepsilon$ so ist der Würfel zulässig.

Datenstruktur & Nachbarn(W)

Es wird eine möglichst einfache Datenstruktur verwendet, welche sich leicht für beliebige Dimensionen verallgemeinern lässt. Sei hierzu W_0 der initiale Würfel im $\mathbb{R}^q$. Er wird durch das q-Tupel $(0,\ldots,0)^T \in \mathbb{Z}^q$ repräsentiert. Die beiden Nachbarwürfel entlang der ersten Dimension sind dann durch $(-1,0,\ldots,0)^T \in \mathbb{Z}^q$ und $(1,0,\ldots,0)^T \in \mathbb{Z}^q$ gegeben. Die Menge aller Nachbarn eines beliebigen Würfels W mit dem Tupel $\alpha = (\alpha_1,\ldots,\alpha_q)^T$ ist gegeben durch

$$\text{Nachbarn}(W) := \left\{ \beta \in \mathbb{Z}^q : \sum_{i=1}^{q} |\alpha_i - \beta_i| = 1 \right\} .$$

Die Anwendung auf eine Menge von Würfeln $\mathcal{W}$ ist durch

$$\text{Nachbarn}(\mathcal{W}) = \bigcup_{W \in \mathcal{W}} \text{Nachbarn}(W)$$

definiert. Auch die Eckpunkte von W lassen sich schnell aus den Eckpunkten von W_0 bestimmen. Sei $e \in \mathcal{G}$ eine Ecke von W und $e_0 \in \mathcal{G}$ die analog positionierte[1] Ecke von W_0. Es gilt $e = e_0 + \omega\alpha$. Auch die Bestimmung des Tupels $\alpha = (\alpha_1,\ldots,\alpha_q)^T$ eines einschließenden Würfels W zu einem gegebenen Vektor $k \in \mathbb{R}^q$ ist einfach durchzuführen. Es sei W_0 so gewählt, dass sich k^0 im Mittelpunkt von W_0 befindet. Mit kaufmännischem Runden kann das zu W gehörende Tupel durch $\alpha = \text{round}(\omega^{-1}(k - k^0))$ bestimmt werden.

Schnelltest(W)

Die zu Beginn des Kapitels erwähnte Beschränktheit von $\mathcal{G}$ wird algorithmisch durch lineare Gleichungs- und Ungleichungs- sowie nichtlineare Gleichungsnebenbedingungen umgesetzt. Es folgt

$$\mathcal{G} := \{k \in \mathbb{R}^q : A_e k = b_e \wedge \; A_i k \leq b_i \wedge \; g(k) = 0\}$$

mit Matrizen A_e, A_i und Vektoren b_e, b_i geeigneter Dimension, sowie einer gegebenenfalls vektorwertigen Funktion $g(k)$. Der Schnelltest eines Würfels W bestimmt durch einfache Matrix-Vektor-Operationen, ob ein Punkt innerhalb des Würfels existiert, der sowohl die linearen Ungleichungs- als auch Gleichungsnebenbedingungen erfüllt.

1. Die Ungleichungsnebenbedingungen lassen sich durch Auswertung von $A_i v \leq b_i$ für die Eckpunkte $v \in \mathbb{R}^q$ des Würfels (als dessen extremale Punkte) überprüfen.
2. Weiter werden die Gleichungsnebenbedingungen ganz analog durch Auswerten von $A_e v \leq b_e$ und $A_e v \geq b_e$ für alle Eckpunkte v geprüft. Werden beide Ungleichungen von mindestens einer Ecke (nicht notwendigerweise die selbe) erfüllt, so existiert im Würfel mindestens ein Vektor v_e, sodass $A_e v_e = b_e$ gilt.

Der Schnelltest ist bestanden, wenn beide "Untertests" bestanden sind.

Zulässigkeitstest(W, ε)

Der Zulässigkeitstest eines Würfels umfasst die Durchführung einer Minimierung der Funktion $f(k)$ aus Gleichung (6.1) innerhalb der Grenzen des Würfels W. Um ein Einhal-

[1] Im zweidimensionalen Fall wäre unter "analog positioniert" zu verstehen, dass e und e_0 beispielsweise jeweils der *linken oberen* Ecke zugeordnet werden.

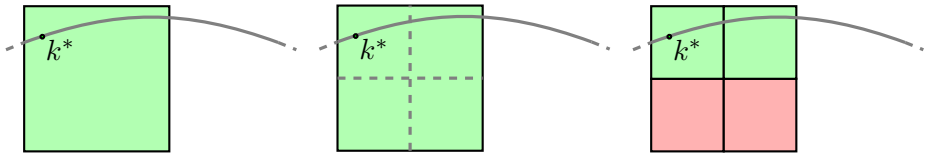


Abbildung 6.3.: Verfeinerung eines Würfels (links) durch Unterteilung in gleichgroße Teilwürfel (mitte) und anschließendem Testen auf Zulässigkeit (rechts).

ten der Nebenbedingungen zu gewährleisten wird statt $f(k)$ die modifizierte Zielfunktion

$$f_Q(k) := f(k) + \|A_e k - b_e\|_2^2 + \|\max(A_i k - b_i, 0)\|_2^2 + \|g(k)\|_2^2 \tag{6.3}$$

genutzt. Der Einfachheit halber erfolgt die Formulierung in (6.3) mittels absoluter Fehlerterme. Für die Berechnung werden relative Fehlerterme empfohlen. Die Restriktion der Optimierung auf den Würfel wird durch zusätzliche Ungleichungsnebenbedingungen umgesetzt. Als Startvektor der Optimierung $k^{(0)}$ wird der Mittelpunkt des Würfels W genutzt. Ist der Funktionswert $f_Q(k^*)$ des optimierten Vektors kleiner oder gleich einer vorgegebenen Fehlertoleranz ε, so ist der Würfel W zulässig. Die finale Iterierte k^* der Minimierung wird für jeden Würfel abgespeichert. Der beschriebene Entscheidungsprozess ist in Abbildung 6.2 vereinfacht dargestellt. Zur Verbesserung der Robustheit des Algorithmus kann die Optimierung gegebenenfalls unter Verwendung unterschiedlicher Startvektoren wiederholt werden. Es bieten sich beispielsweise die Ecken des jeweiligen Würfels als extremale Punkte an. Die Optimierung kann in Matlab mit der Routine *lsqnonlin* durchgeführt werden.

6.2.1. Gitterverfeinerung

Ausgehend von einer Menge von Würfeln $\mathcal{W}$, welche mit dem Algorithmus 1 bestimmt wurde, kann leicht eine Menge von Würfeln $\mathcal{W}^+$ zu einer kleineren Würfelkantenlänge berechnet werden. Die Idee besteht darin jeden Würfel beispielsweise durch Halbieren der Seitenkanten in Teilwürfel zu zerlegen und diese wiederum mit den im vorherigen Abschnitt vorgestellten Methoden auf Zulässigkeit zu testen. Das Vorgehen ist in Algorithmus 2 beschrieben und eine vereinfachte Darstellung der Vorgehensweise ist in Abbildung 6.3 gezeigt.

Darüber hinaus kann eine adaptive Strategie verwendet werden, um nur solche Würfel zu verfeinern, die in Kombination mit ihren Nachbarn keine ausreichende lineare Approximation zwischen den jeweiligen finalen Iterierten k^* ermöglichen. Hierzu werden bei der Berechnung von "Unterteilung(W)" die folgenden Schritte zusätzliche durchgeführt:

1. Bestimme $N_{\mathcal{W}}$ =Nachbarn(W) $\cap$ $\mathcal{W}$.
2. Sei $k^*(W)$ die einem Würfel W zugeordnete finale Iterierte der Minimierung im Zulässigkeitstest. Für alle $N \in N_{\mathcal{W}}$ wird der Mittelpunkt $k^{\text{test}} = (k^*(W)+k^*(N))/2$ der optimierten Vektoren von W und N bestimmt. Gilt $f_Q(k^{\text{test}}) \leq \varepsilon$ hat der Nachbarwürfel N den Test bestanden.
3. Bestehen alle Nachbarn von W den in Punkt 2 beschriebenen Test wird angenommen, dass eine lineare Approximation zwischen W und seinen Nachbarn hinreichend gut ist und es wird von einer weiteren Verfeinerung von W abgesehen. Andernfalls erfolgt die Unterteilung von W in Teilwürfel.

Algorithmus 2 Verfeinerungsschritt

Input: Einschließung von $\mathcal{N}$ durch die Menge von Würfeln $\mathcal{W}$, Anzahl der Verfeinerungsschritte $\nu \in \mathbb{N}$, Fehlertoleranz $\varepsilon \geq 0$
Output: Einschließung $\mathcal{W}^+$ von $\mathcal{N}$ auf feinerem Gitter

for $i = 1, \ldots, \nu$ **do**
  $\mathcal{W}^+ = \emptyset$
  for all $W \in \mathcal{W}$ **do**
    for all $W^u \in$ Unterteilung(W) **do**
      Bestimme k^* durch Minimierung von f_Q mit einem Startvektor innerhalb von W^u
      if $f_Q(k^*) \leq \varepsilon$ **then**
        $\mathcal{W}^+ = \mathcal{W}^+ \cup \{W^u\}$
      end if
    end for
  end for
  $\mathcal{W} = \mathcal{W}^+$
end for

6.2.2. Parallelisierung

Der zeitaufwändige Teil der Algorithmen 1 und 2 besteht in der Durchführung der Optimierung innerhalb des *Zulässigkeitstests*. Dieser Test muss innerhalb einer Iteration nicht nur für einen Würfel, sondern für eine Menge von Würfeln durchgeführt werden. Hieraus ergibt sich also eine Menge von Optimierungsproblemen, welche unabhängig von einander gelöst werden können. Da moderne Computer häufig über mehr als einen CPU-Kern verfügen, können beispielsweise die *Parallel Computing Toolbox* von Matlab oder die *pthreads*-Bibliothek in C effektiv zur Parallelisierung dieser Aufgaben und somit zur Reduzierung der Rechenzeit eingesetzt werden. In der bereitgestellten Implementierung des Würfeleinschließungsalgorithmus des Autors in Matlab ist die parallele Nutzung mehrere CPU-Kerne berücksichtigt.

6.3. Numerische Beispiele

In den folgenden Abschnitten wird der Würfeleinschließungsalgorithmus anhand verschiedener Modellprobleme bezüglich der Approximationsgüte gegenüber dem Grid Search Algorithmus (Abschnitt 6.3.1), der Approximationsgüte für (in diesem Kontext) komplexe Mengen $\mathcal{N}$ (Abschnitt 6.3.2) und der Verwendung von Parallelisierungs- und Gitterverfeinerungsstrategien (Abschnitt 6.3.3) untersucht.

6.3.1. Würfeleinschließung vs. Grid Search

Das Ziel des folgenden Modellproblems ist die Approximation einer Höhenlinie unter Verwendung des Matlab-Beispieldatensatz *peaks(100)*. Es ist eine Diskretisierung von $\mathcal{G} = [-3, 3] \times [-3, 3]$ durch Unterteilung der x- und y-Achse mit jeweils 100 äquidistanten Gitterpunkten gegeben. Zu jedem Gitterpunkt (x, y) ist außerdem ein Höhenwert durch den Datensatz definiert. Mit $z(x, y)$ sei der durch lineare Interpolation ermittelte

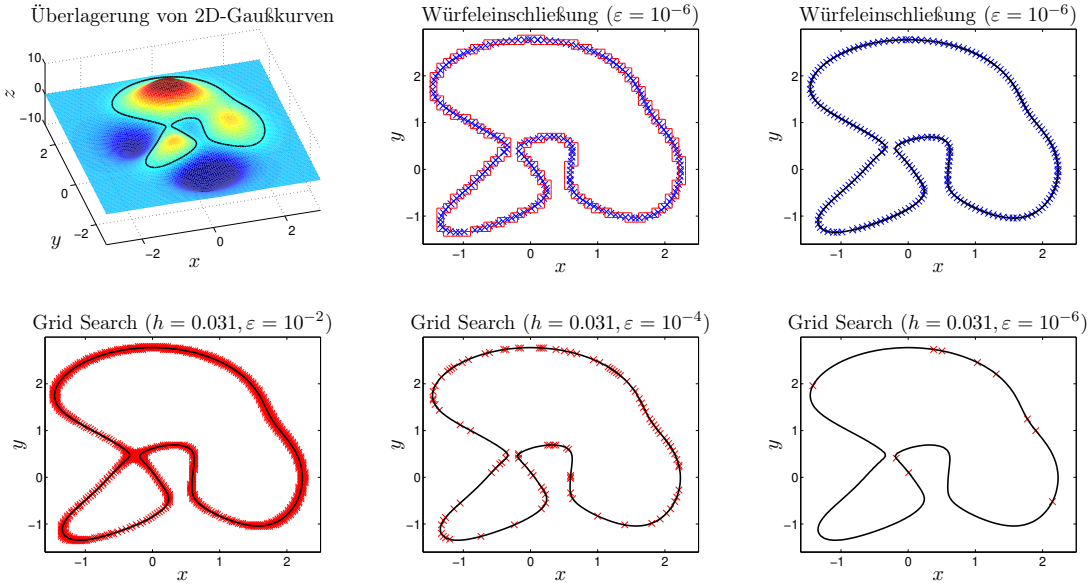


Abbildung 6.4.: Anwendung von Grid Search und der Würfeleinschließung auf den Matlab-Beispieldatensatzes *peaks*(100) zur Berechnung der Höhenlinie für $z = 0.75$. Der Datensatz und die gesuchte Höhenlinie sind in der oberen linken Grafik zu sehen. Die beiden anderen Grafiken in der oberen Reihe zeigen die Approximation der Höhenlinie durch den Würfeleinschließungsalgorithmus als blaue Kreuze. Es wurden eine initiale Kantenlänge von 0.2, ein Verfeinerungsschritt und eine Fehlertoleranz von $\varepsilon = 10^{-6}$ verwendet. Die Vereinigung der einschließenden Würfel ist in rot dargestellt. Die untere Reihe zeigt die Ergebnisse der Anwendung des Grid Search Algorithmus zu verschiedenen Fehlertoleranzen $\varepsilon \in \{10^{-2}, 10^{-4}, 10^{-6}\}$.

Höhenwert für ein beliebiges $(x, y) \in \mathcal{G}$ definiert. Die oben links in der Abbildung 6.4 gezeigte Höhenlinie für $z = 0.75$ (schwarze Linie) ist nun durch den Würfeleinschließungsalgorithmus und Grid Search zu approximieren. Die zugehörige Niveaumenge $\mathcal{N}$ ist durch die Funktion $f(x, y) = z(x, y) - 0.75$ definiert.

In Abbildung 6.4 ist links oben der Datensatz *peaks(100)* veranschaulicht. Weiter ist in dieser und den anderen Grafiken der Abbildung die zu approximierende Höhenlinien als schwarze Linie eingezeichnet. Das Vorgehen zur Approximation von $\mathcal{N}$ mittels Würfeleinschließungsalgorithmus wird nun beschrieben. Als initialer Punkt wird $(x_0, y_0) = (-0.3934, 0.371)^T$ gewählt. Die Kantenlänge ω der Würfel beträgt zunächst 0.2, wobei ein Verfeinerungsschritt durchgeführt und somit eine finale Kantenlänge von 0.1 erreicht wird. Die Optimierung innerhalb des *Zulässigkeitstests* eines jeden Würfels wird abgebrochen, wenn $f_Q(x, y)$ die Fehlertoleranz von $\varepsilon = 10^{-6}$ unterschreitet. In der oberen mittleren Grafik ist die ermittelte Approximation (blaue Kreuze) und die Menge der einschließenden Würfel (rote Umrandung) zu sehen. Die obere rechte Grafik zeigt die jeweiligen optimierten Punkte (x^*, y^*) der einschließenden Würfel. Die Übereinstimmung mit der zu approximierende Höhenlinien in schwarz sowie deren gleichmäßige Verteilung entlang dieser Linie sind gut zu erkennen. Die Rechenzeit beträgt $12.91s$, wobei zwecks Vergleichbarkeit auf Parallelisierung verzichtet wird. In Abbildung 6.5 sind ausgewählte Iterationsschritte sowie ein Verfeinerungsschritt des Würfeleinschließungsalgorithmus gezeigt.

Auch der Grid Search Algorithmus kann zur Ermittlung einer Approximation der Höhenlinie genutzt werden. Um einen sinnvollen Vergleich zu ermöglichen, wurde $h = 0.031$ als Gitterweite gewählt. Die Rechenzeit weicht dann mit $13.02s$ nur gering von der des

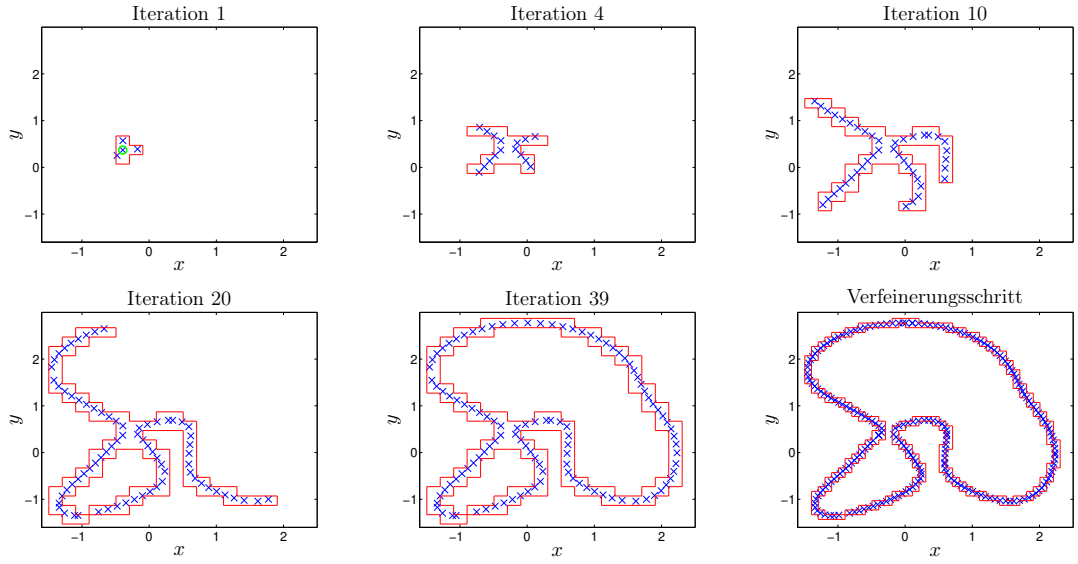


Abbildung 6.5.: Darstellung der Menge der Würfel zu verschiedenen Iterationsschritten sowie einem anschließenden Verfeinerungsschritt für das Modellproblem aus Abschnitt 6.3.1. Im ersten Iterationsschritt ist zusätzlich der initiale Punkt $(-0.3934, 0.371)^T$ als grüner Kreis dargestellt.

Würfeleinschließungsalgorithmus ab. Die unteren Grafiken der Abbildung 6.4 zeigen die Ergebnisse zur unterschiedlichen Fehlertoleranzen ε. Es ist leicht zu sehen, dass eine hohe Fehlertoleranz ε gewählt werden muss um die gesamte Höhenlinie zu erfassen. Die Approximation mit $\varepsilon = 10^{-6}$, wie beim Würfeleinschließungsalgorithmus, ist nicht mehr erfolgreich.

Es kann also zusammengefasst werden, dass durch den Würfeleinschließungsalgorithmus im Gegensatz zum Grid Search Algorithmus bei gleichem zeitlichen Aufwand eine bessere Approximation der Höhenlinie erreicht wird. Dies bezieht sich auf die Verteilung der ermittelten Punkte entlang der Höhenlinie und auf die höhere Approximationsgüte in Form deutlich geringerer Fehlertoleranzen ε.

6.3.2. Approximation komplexer Strukturen

Die Möglichkeit, den Würfeleinschließungsalgorithmus für Strukturen verschiedenster Art einsetzen zu können, soll im Folgenden verdeutlicht werden. Exemplarisch werden zwei simulierte Niveaumengen betrachtet: Erstens, eine 3D-Struktur, welche von einer Linie in eine Fläche und wiederum in ein Volumen übergeht. Zweitens, eine Vereinigung der Graphen der Sinus- und Kosinusfunktion.

Die erste zu approximierende Niveaumenge $\mathcal{N}_1$ wird durch die Funktion $f_1 : \mathcal{G}_1 \to \mathbb{R}$ mit

$$f_1(x,y,z) := \begin{cases} \|(x,y)\|_2 & \text{für } z \leq 1 \\ \|(\max(|x| - z + 1, 0), y)\|_2 & \text{für } 1 < z \leq 2 \\ \|(\max(|x| - 1, 0), \max(|y| - z + 2, 0))\|_2 & \text{für } 2 < z \end{cases}$$

und $\mathcal{G}_1 = [-1,1] \times [-1,1] \times [0,3]$ definiert. Als initialer Punkt wird $(0,0,0)^T$ gewählt. Die Kantenlänge der Würfel beträgt zunächst 0.3, wobei ein Verfeinerungsschritt durchgeführt und somit eine finale Kantenlänge von 0.15 erreicht wird. Die Fehlertoleranz ε beträgt 10^{-8}. Die Ergebnisse sind in den oberen zwei Grafiken der Abbildung 6.6 zu

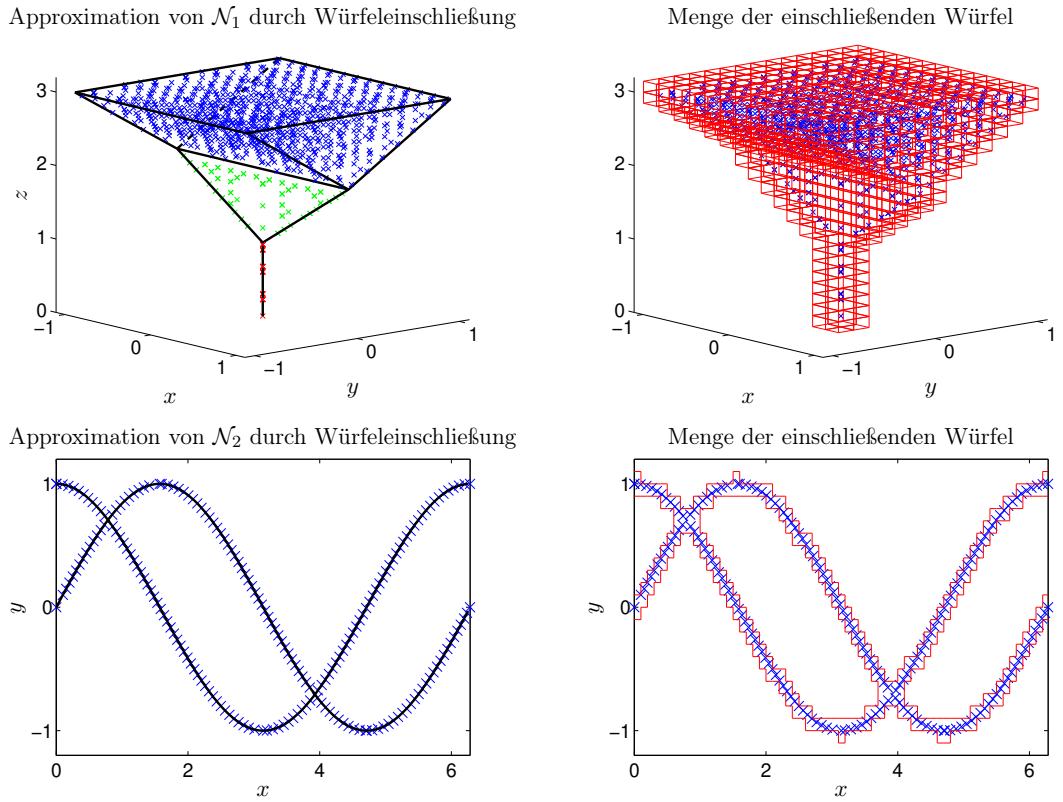


Abbildung 6.6.: Anwendung des Würfeleinschließungsalgorithmus auf zwei Modellprobleme. Oben: Beide Grafiken zeigen die Approximation (Kreuze) einer 3D-Struktur, welche in einen ein-, zwei- und dreidimensionalen Bereich unterteilt ist. In der linken Grafik wurde eine zusätzliche farbliche Unterscheidung vorgenommen sowie die Kanten der zu approximierenden Struktur eingezeichnet. Die rechte Grafik zeigt die einschließende Menge der Würfel in rot. Unten: Die zu approximierende Struktur besteht aus der Vereinigung der Sinus- und der Kosinusfunktion auf dem Intervall $[0, 2\pi]$. Die analytische Lösung ist als schwarze Linie in der linken Grafik eingezeichnet. Auch hier zeigt die rechte Grafik die einschließende Menge der Würfel in rot.

sehen. Für den gewählten initialen Punkt erfolgt die Ausbreitung der Würfel von "von unten nach oben". Die dabei auftretenden Übergänge von ein- zu zweidimensionalen und zwei- zu dreidimensionalen Strukturen stellen keine Probleme dar. In der rechten Grafik ist zu erkennen, dass eine komplette Einschließung der zu approximierenden Struktur durch Würfel erreicht wird.

Die zweite zu approximierende Niveaumenge $\mathcal{N}_2$ wird durch die Funktion $f_2 : \mathcal{G}_2 \to \mathbb{R}$ mit

$$f_2(x, y) := \min(|y - \sin(x)|, |y - \cos(x)|)$$

und $\mathcal{G}_2 = [0, 2\pi] \times [-1, 1]$ definiert. Als initialer Punkt wird $(0, 0)^T$ gewählt. Die Kantenlänge der Würfel beträgt zunächst 0.2, wobei ein Verfeinerungsschritt durchgeführt und somit eine finale Kantenlänge von 0.1 erreicht wird. Die Fehlertoleranz ε beträgt 10^{-8}. Die Überlagerung von Sinus- und Kosinusfunktion führt dazu, dass die zu approximierende Struktur ein Loch und Verzweigungen aufweist. Die Ergebnisse sind in den unteren zwei Grafiken der Abbildung 6.6 zu sehen. In den kritischen Stellen, den Schnittpunkten von Sinus- und Kosinusfunktion, breiten sich die Würfel in alle relevanten Richtungen aus. Somit kann auch für dieses Beispiel eine komplette Einschließung der zu approximierenden Struktur durch Würfel erreicht werden.

6.3.3. Reduktion der Rechenzeit durch Parallelisierung und adaptive Gitterverfeinerung

Durch parallelisierte Ausführung von Teilen der Algorithmen 1 und 2 sowie die adaptive Gitterverfeinerung kann eine Reduktion der Rechenzeit erreicht werden.

In Abbildung 6.7 wird für zwei Computersysteme die Auswirkung der Parallelisierung auf das Modellproblem des kinetischen Modells $A \Leftrightarrow B \Leftrightarrow C$ aus Abschnitt 4.1.2 veranschaulicht. Die sehr gute Skalierbarkeit des Würfeleinschließungsalgorithmus ist anhand der Rechenzeiten in Abhängigkeit der verwendeten CPU-Kerne/Threads in den folgenden Tabellen 6.1 erkennbar:

		Anzahl verwendeter Threads (Intel i7-4790)			
		1	2	4	6
϶	0.3	140.9s	78.3s	54.3s	42.7s
	0.5	52.9s	29.9s	18.6s	16.8s
	0.7	29.3s	16.1s	10.3s	10.1s
	0.9	18.6s	10.3s	6.7s	6.4s

		Anzahl verwendeter Threads (Intel Xeon Gold 6144)				
		1	2	4	8	16
϶	0.3	91.0s	48.1s	27.8s	15.6s	11.1s
	0.5	35.1s	18.8s	11.0s	6.9s	5.2s
	0.7	18.6s	10.3s	6.5s	4.1s	3.3s
	0.9	12.0s	7.0s	4.3s	3.2s	2.7s

Tabelle 6.1.: Rechenzeit zur Lösung des Modellproblems $\mathcal{X} \rightleftharpoons \mathcal{Y} \rightleftharpoons \mathcal{Z}$ aus Abschnitt 4.1.2 in Abhängigkeit von Würfelkantenlänge ω und der Anzahl der verwendeten Threads für zwei Computersysteme. Auf Verfeinerungsschritte wird verzichtet. Für beide Systeme und einen bis vier Threads ist eine gute Skalierbarkeit des Problems festzustellen, da eine Verdoppelung der Threadanzahl in etwa einer Halbierung der Rechenzeit entspricht. Für das i7-System mit 6 Threads ist nur noch eine geringfügige Verbesserung zu erkennen, da nur 4 physische CPU-Kerne vorhanden sind. Für das Xeon-System mit 8 und 16 Threads nimmt die Rechenzeit ab, allerdings nur noch in einem geringeren Maße als von 1 zu 2 oder 2 zu 4 Threads. Die Skalierbarkeit lässt im Allgemeinen nach, wenn die Anzahl der verwendeten Threads nahe an der Anzahl der physischen CPU-Kerne liegt.

Hierbei ist anzumerken, dass der Würfeleinschließungsalgorithmus zur Approximation von Niveaumengen $\mathcal{N}$ mit einfachen Strukturen wie Linien oder Kurven (vergleiche die kinetischen Modelle $\mathcal{X} \longrightarrow \mathcal{Y} \rightleftharpoons \mathcal{Z}$ oder $\mathcal{X} \rightleftharpoons \mathcal{Y} \longrightarrow \mathcal{Z}$ in Abschnitt 4.1.2) nur wenig von einer hohen Kernanzahl profitiert. Dies ist dadurch begründet, dass dann in einer Iterationen des Würfeleinschließungsalgorithmus immer nur wenige Würfel getestet werden. Beispielsweise werden für eine Niveaumenge $\mathcal{N}$ mit Linienstruktur in jeder Iteration nur zwei Würfel auf Zulässigkeit getestet. Für die Approximation "komplexerer" Niveaumengen (zum Beispiel Flächen, Volumina, etc.) ist die Anzahl zu prüfender Würfel typischerweise nach nur wenigen Iteration deutlich größer als die Anzahl der verfügbaren physischen CPU-Kerne. Damit kann eine hohe Auslastung aller zur Verfügung stehenden Kerne und eine gute Skalierbarkeit des Problems mit der Kernanzahl erwartet werden.

Anhand des selben Modellproblems kann auch der Einfluss der adaptiven Gitterverfeinerung auf die Rechenzeit untersucht werden. Hierzu wird der Würfeleinschließungsalgorithmus mit einer Fehlertoleranz $\varepsilon = 10^{-9}$, $\nu = 4$ Verfeinerungsschritten mit initialer

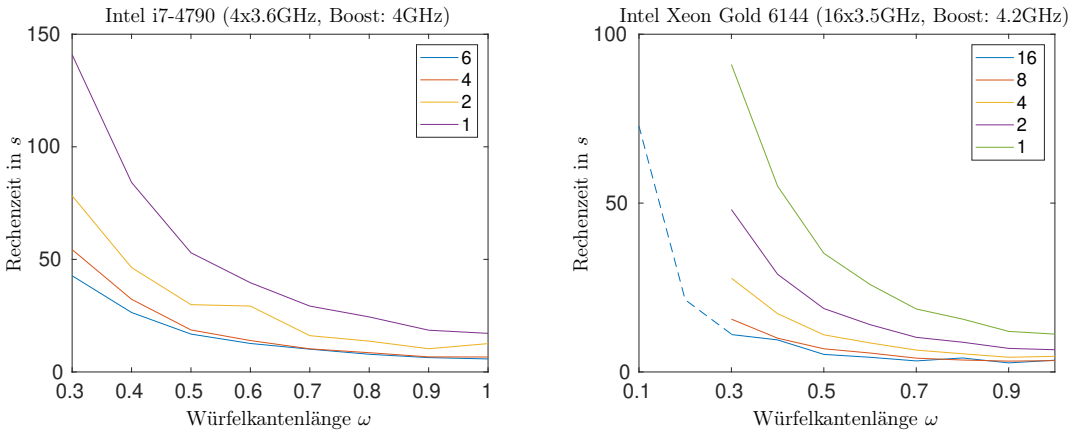


Abbildung 6.7.: Vergleich der Rechenzeit zur Lösung des Modellproblems $\mathcal{X} \rightleftharpoons \mathcal{Y} \rightleftharpoons \mathcal{Z}$ aus Abschnitt 4.1.2 in Abhängigkeit von Würfelkantenlänge ω und der Anzahl der verwendeten Threads (siehe Legende). Die Rechenzeit sinkt wie erwartet bei steigender Anzahl verwendeter Threads. Die Skalierbarkeit ist besonders im Bereich von einem bis vier Kernen gegeben, das heißt eine Verdopplung der Kerne entspricht in etwa einer Halbierung der Rechenzeit. In Bezug auf den Intel i7-4790 ist über 4 Threads hinweg keine starke Verbesserung zu erwarten, da die CPU nur 4 physische Kerne besitzt.

Kantenlänge $\omega = 1$ und parallelisierter Rechnung unter Nutzung von maximal 4 Threads genutzt. Die adaptive Strategie benötigt mit 70.1s im Gegensatz zur nichtadaptiven Variante mit 262s nur etwa 1/4 der Zeit. Ein Vergleich der Ergebnisse ist in Abbildung 6.8 zu sehen. Die verschiedenen Verfeinerungsstufen sind in der rechten Grafik durch unterschiedliche Farben gekennzeichnet. Darüber hinaus sind in Abbildung 6.9 die Approximationen der Niveaumengen zu den Modellproblemen aus den Abschnitten 6.3.1 und 6.3.2 unter Verwendung von $\nu = 3$ adaptiven Verfeinerungsschritten dargestellt. Die Zeitersparnis beträgt 55% für die Überlagerung der Sinus- und Kosinusfunktion sowie 25% für das Modellproblem zum Datensatz *peaks*(100).

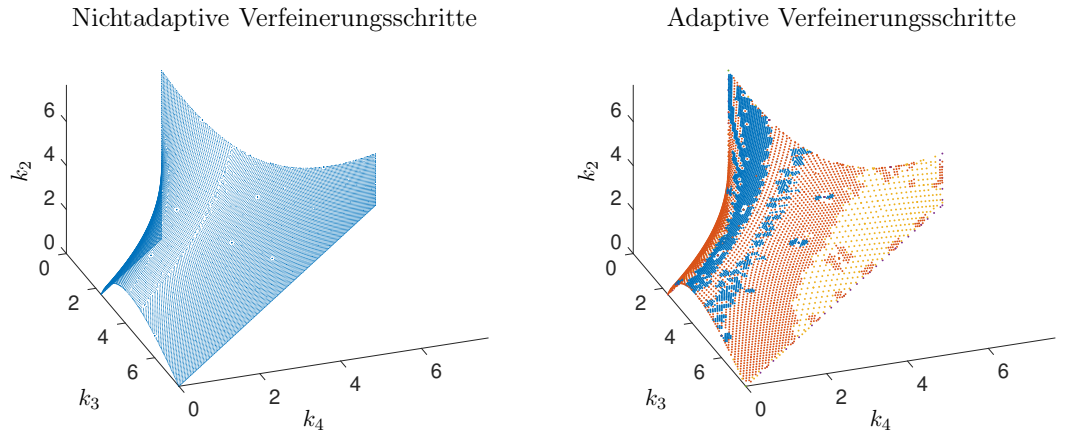


Abbildung 6.8.: Approximationen der Menge $\mathcal{K}$ des Modellproblems $\mathcal{X} \rightleftharpoons \mathcal{Y} \rightleftharpoons \mathcal{Z}$ aus Abschnitt 4.1.2 durch den Würfeleinschließungsalgorithmus unter Verwendung von vier nichtadaptiven (links) und adaptiven (rechts) Verfeinerungsschritten. In der rechten Grafik sind die Würfelrepräsentanten der unterschiedlichen Verfeinerungsstufen farblich markiert: grün (grob) → lila → gelb → rot → blau (fein).

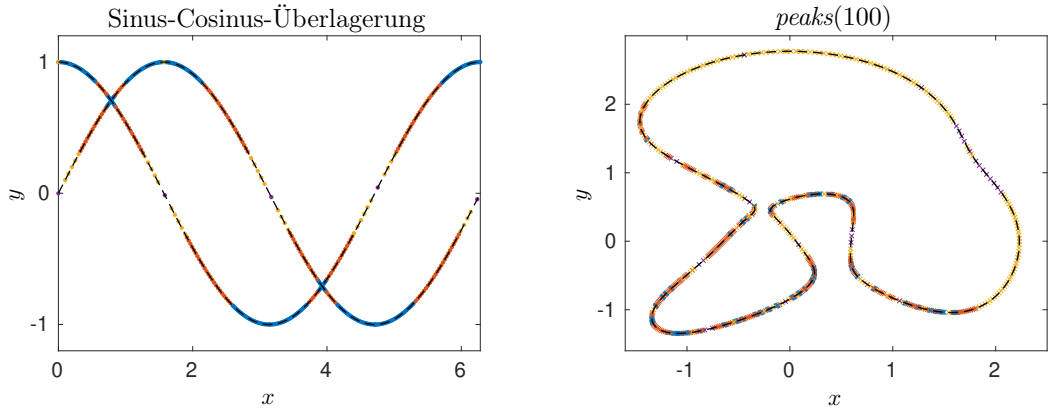


Abbildung 6.9.: Approximationen der Niveaumengen für zwei Beispiele aus den Abschnitten 6.3.1 und 6.3.2 durch den Würfeleinschließungsalgorithmus unter Verwendung von drei adaptiven Verfeinerungsschritten. Die Würfelrepräsentanten der unterschiedlichen Verfeinerungsstufen sind farblich markiert: lila (grob) $\rightarrow$ gelb $\rightarrow$ rot $\rightarrow$ blau (fein). In schwarz sind links die Verläufe von $\sin(x)$ und $\cos(x)$ sowie rechts die Höhenlinie für $z = 0.75$ des Matlab-Datensatzes *peaks*(100) eingezeichnet.

7. Anwendungen in der Chemometrie

In diesem Kapitel werden exemplarisch drei Matrixfaktorisierungsprobleme unter Verwendung der Regularisierung mittels eines kinetischen Modells präsentiert. Es werden eine Rhodiumclusterbildung und eine Katalysatorpräformierung aus dem Bereich der Hydroformylierung sowie eine Isomerisierung aus dem Bereich der Photochemie betrachtet. Eine Übersicht der jeweiligen Eigenschaften und Ergebnisse ist in der folgenden Tabelle gegeben:

Themengebiet	Kinetik	Datensatz	$D \to C, S$	$\mathcal{K}$	$\mathcal{K}^+$	$\mathcal{K}_\varepsilon$	$\mathcal{K}^+_{\varepsilon,\theta}$
Rh-Clusterbildung (Abschnitt 7.2)	2. Ord.	FTIR	✓	✗	✗	✗	✗
Photokinetik (Abschnitt 7.3.1)	1. Ord.	UV/Vis (Multiset)	✓	✓	✓	✓	✓
Präformierung (Abschnitt 7.3.2)	1. Ord.	FTIR	✓	✓	✓	✓	✓

7.1. Datenvorbehandlung

Typischerweise beinhaltet eine gemessene Spektrenfolge D eine Reihe von Störungen, wie beispielsweise Rauschen oder Grundlinienfehler. Durch die Nutzung der Niedrigrangapproximation von D kann nicht selten eine Reduktion des Rauschens erreicht werden. Als grobe Faustregel sollte die Intensität des Rauschens nicht mehr als $5 - 10\%$ der Intensität der zu analysierenden Peaks ausmachen. Wichtiger ist die Korrektur von Grundlinienfehlern. Diese werden beispielsweise durch systematische Messfehler, den Abzug von Hintergrundspektren oder auch die Überlagerung der zu analysierenden Peaks mit Peaks weiterer Komponenten hervorgerufen. Algorithmen zur Korrektur sind unter anderem in [45, 97] zu finden. Wegen der häufig sehr guten Ergebnisse sei die Modellierung einer fehlerhaften Grundlinie durch Polynome niedrigen Grades (≤ 5) explizit erwähnt. Dieser Ansatz wurde zur Vorbehandlung aller FTIR-Datensätze dieses Kapitels genutzt. Für die UV/Vis Spektrenfolgen in dieser Arbeit sind keine derartigen Vorbehandlungsschritte nötig.

7.2. Regularisierte Matrixfaktorisierung für eine Rhodiumdimerbildung

Es wird nun der in Abschnitt 3.1 vorgestellte Hard-Modell-Ansatz auf einen FTIR-spektroskopischen Datensatz angewandt. Der Fokus liegt also auf der Bestimmung einer regularisierten Matrixfaktorisierung mit simultaner Parametrierung eines kinetischen Modells.

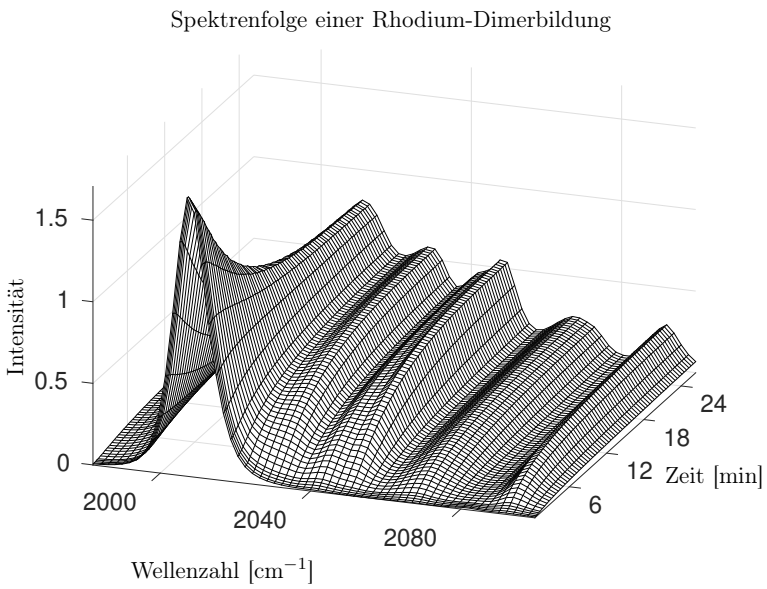
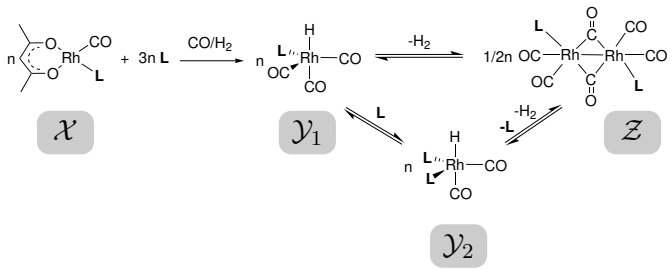


Abbildung 7.1.: Spektrenfolge zur Untersuchung der Rhodiumclusterbildung während einer Katalysatorpräformierung. Eine Datenvorbehandlung in Form einer Grundlinienkorrektur sowie dem Entfernen der Wellenzahlbereichen ohne signifikante Absorptionen an den Rändern wurde durchgeführt.

Es wird eine FTIR-Spektrenfolge zur Analyse der Bildung eines Rhodium-Dimerkomplexes während einer Katalysatorpräformierung betrachtet. Im Detail ist die Reaktionsformel durch



(7.1)

gegeben. Dieses Reaktionssystem ist in das chemischen Themengebiet der Hydroformylierung einzuordnen. Die Versuchsdurchführung erfolgte durch Dr. Christoph Kubis am LIKAT in Rostock. Für die Spektrenfolge wurde eine Grundlinienkorrektur durchgeführt und es wurden Wellenzahlbereiche ohne relevante Informationen entfernt. Dies resultiert in der in Abbildung 7.1 gezeigten Spektrenfolge $D \in \mathbb{R}^{m \times n}$ mit $m = 1495$ Spektren innerhalb des Zeitbereichs $[0.44\text{min}, 26.63\text{min}]$ zu jeweils $n = 493$ Wellenzahlen im Intervall $[1982\text{cm}^{-1},\ 2101\text{cm}^{-1}]$. Zusammengefasst seien die folgenden Systemeigenschaften genannt:

- Spezies: Präkatalysator $\mathcal{X}$=Rh(acac)(CO)L, zwei Katalysatoren $\mathcal{Y}_1$=$\text{HRh(CO)}_3\text{L}$ und $\mathcal{Y}_2$=$\text{HRh(CO)}_2\text{L}_2$ sowie der dinukleare Komplex $\mathcal{Z}$=$\text{Rh}_2\text{(CO)}_6\text{L}_2$, siehe Gleichung (7.1). Die Komponenten $\mathcal{Y}_1$ und $\mathcal{Y}_2$ befinden sich in einem schnellen Gleichgewicht und können nicht getrennt von einander beobachtet werden. Sie werden zu einer Komponente $\mathcal{Y}$ zusammengefasst. Die $\mathcal{X}, \mathcal{Y}$ und $\mathcal{Z}$ zugeordneten Konzentrationsverläufe seien mit $x(t), y(t)$ und $z(t)$ bezeichnet.
- Startkonzentrationen: $c_0 = (x(0), y(0), z(0)) = (20, 0, 0)$ in mmol/ℓ.

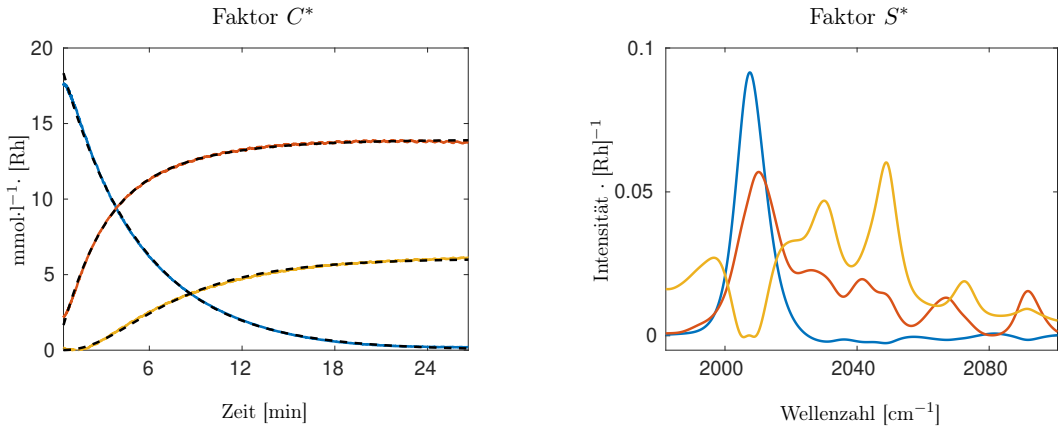


Abbildung 7.2.: Ergebnis der Zerlegung des in Abbildung 7.1 gezeigte Datensatzes. Links sind die Konzentrationsverläufe zu $\mathcal{X}$ (blau), $\mathcal{Y}$ (rot), $\mathcal{Z}$ (gelb) und das mit k^* parametrierte kinetische Modell als gestrichelte, schwarze Linien dargestellt. Die ermittelten Spektren sind mit der gleichen Farbcodierung rechts zu sehen.

- Kinetik: $\mathcal{X} \xrightarrow{k_1} \mathcal{Y},\ 2\mathcal{Y} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathcal{Z}$
- Differentialgleichungssystem:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{pmatrix} -k_1 & 0 & 0 \\ k_1 & -2k_2 & 2k_{-2} \\ 0 & k_2 & -k_{-2} \end{pmatrix} \begin{pmatrix} x(t) \\ y(t)^2 \\ z(t) \end{pmatrix}$$

Zur Optimierung der Zielfunktion (3.6) aus Abschnitt 3.1 mit den Parametern $\gamma_1 = \gamma_2 = \gamma_3 = 1$ und $\delta_1 = 0.1$ wird die Matlab-Routine *lsqnonlin* mit den Parametern $TolX = TolFun = 10^{-10}$ verwendet. Die Reaktionsgeschwindigkeiten wurden mit

$$k^{(0)} = \left(0.001\text{s}^{-1},\ 0.01\text{s}^{-1}\ell\ \text{mmol}^{-1},\ 0.01\text{s}^{-1}\right)$$

initialisiert, womit sich die optimierten Parameter

$$k^* = \left(3.23 \cdot 10^{-3}\text{s}^{-1}, 9.44 \cdot 10^{-5}\text{s}^{-1}\ell\ \text{mmol}^{-1}, 6.05 \cdot 10^{-3}\text{s}^{-1}\right)$$

ergeben. Die resultierenden Konzentrations- sowie Spektrenverläufe sind in Abbildung 7.2 zu sehen. Zur besseren chemischen Interpretation und Übersicht der grafischen Darstellung wurde eine Skalierung basierend auf der Anzahl der Rhodiumatome in den Spezies eingeführt. Das Spektrum des Rh-Dimers (gelb) wird durch 2 dividiert und das zugehörige Konzentrationsprofil mit 2 multipliziert. Die Residuen zu den einzelnen Summanden der verwendeten Zielfunktion lauten

$$\frac{\|D - C^*(S^*)^T\|_F}{\|D\|_F} = 0.0046\ , \qquad \frac{\|C^* - C^{\text{dgl}}(k^*)\|_F}{\|C^*\|_F} = 0.00083\ ,$$
$$\frac{\|\min(C^*, 0)\|_F}{\|C^*\|_F} = 8.3 \cdot 10^{-5}\ , \qquad \frac{\|\min(S^*, 0)\|_F}{\|S^*\|_F} = 0.018$$

und werden als hinreichend gut eingestuft.

7.3. Mengen zulässige Parameter für (photo-)kinetische Modelle

In den folgenden Abschnitten erfolgt die Analyse der Menge D-konsistenter und D-approximativer Parameter für eine UV/Vis- und eine FTIR-Messserie.

7.3.1. Phototansformation von CMTE

Bei dem vorliegende Reaktionssystem handelt es sich um eine Fallstudie über *cis*-1,2-dicyano-1,2-bis(2,4,5-trimethyl-3-thienyl)ethen (CMTE) aus dem Bereich der Photochemie [30, 31]. Die Daten wurde bereitgestellt durch Cyril Ruckebusch und Olivier Devos der "Université des Sciences et Technologies" in Lille sowie Rémi Métivier der ENS[1] in Cachan. Es werden eine regularisierte Matrixfaktorisierung sowie die Mengen $\mathcal{K}$, $\mathcal{K}^+$, $\mathcal{K}_\varepsilon$ und $\mathcal{K}^+_{\varepsilon,\theta}$ bestimmt. Diese Parametermengen enthalten im Gegensatz zu den bisher betrachteten Kinetiken keine Geschwindigkeitsparameter sondern sogenannte Quantenausbeuten, welche mit ϕ bezeichnet werden.

Spektroskopischer Datensatz 2. Die Messserie setzt sich aus zwei Teildatensätzen zusammen. Hierbei sind alle Versuchsparameter mit Ausnahme der Startkonzentrationen $c_{1,0}$ und $c_{2,0}$ identisch. Die Datenvorbehandlung umfasst lediglich das Entfernen eines Wellenzahlbereichs, welcher Artefakte in Form von starkem Rauschen enthält. Dies wird in analoger Form für beide Teildatensätze durchgeführt. In Abbildung 7.3 sind die vorbehandelten Datensätze $D_1 \in \mathbb{R}^{m_1 \times n}$ und $D_2 \in \mathbb{R}^{m_2 \times n}$ dargestellt. Sie umfassen $m_1 = 103$ und $m_2 = 146$ Spektren im Zeitbereich $[0\text{s}, 702\text{s}]$ beziehungsweise $[0\text{s}, 996\text{s}]$ zu jeweils $n = 621$ Wellenlängen im Bereich $[275\text{nm},\ 440\text{nm}] \cup [501\text{nm},\ 628\text{nm}]$. Die beiden Systeme wurde mit Licht der Wellenlänge 405nm der Intensitäten $I_1 = 4.8 \cdot 10^{-6}$ und $I_2 = 4.5 \cdot 10^{-6}$ in mol $\ell^{-1}\text{s}^{-1}$ angeregt. Weitere Systemeigenschaften seien im Folgenden zusammengefasst:

- Spezies: $\mathcal{X}$=offenes *cis* CMTE-Isomer, $\mathcal{Y}$=geschlossene Kreisform von CMTE und $\mathcal{Z}$=offenes *trans* CMTE-Isomer mit zugeordneten Konzentrationsverläufen $x(t)$, $y(t)$ und $z(t)$.
- Molare Absorptionskoeffizienten: $\varepsilon_\mathcal{X} = 3.1075 \cdot 10^3, \varepsilon_\mathcal{Y} = 1.0862 \cdot 10^3, \varepsilon_\mathcal{Z} = 2.3365 \cdot 10^3$ in mol $\ell^{-1}\text{s}^{-1}$.
- Startkonzentrationen: $c_{1,0} = (5.93 \cdot 10^{-5}, 0, 0)$ und $c_{2,0} = (0, 0, 1.072 \cdot 10^{-4})$ in mol ℓ^{-1}.
- Photokinetisches Modell: $\mathcal{Y} \underset{\phi_1}{\overset{\phi_{-1}}{\rightleftharpoons}} \mathcal{X} \underset{\phi_{-2}}{\overset{\phi_2}{\rightleftharpoons}} \mathcal{Z}$
- Differentialgleichungssystem mit photokinetischem Faktor $F(t)$:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} = F(t) \cdot I \cdot \underbrace{\begin{pmatrix} -\phi_1 - \phi_2 & \phi_{-1} & \phi_{-2} \\ \phi_1 & \phi_{-1} & 0 \\ \phi_2 & 0 & \phi_{-2} \end{pmatrix} \begin{pmatrix} \varepsilon_\mathcal{X} & & \\ & \varepsilon_\mathcal{Y} & \\ & & \varepsilon_\mathcal{Z} \end{pmatrix}}_{M(\phi)} \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$$

[1] École normale supérieure Paris-Saclay

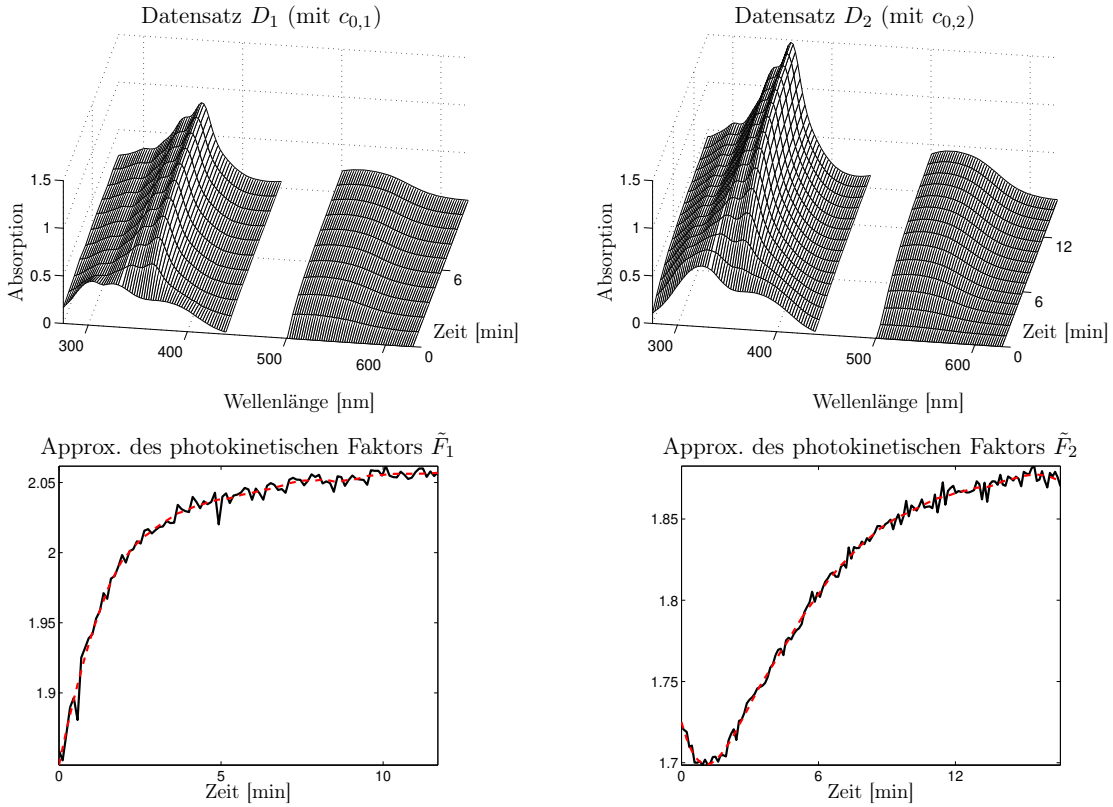


Abbildung 7.3.: Oben sind die zwei UV/Vis-spektroskopischen Messserien aus dem spektroskopischen Datensatz 2 zur Untersuchung der Phototransformation von CMTE dargestellt. Die approximierten Verläufe der photokinetischen Faktoren $\tilde{F}_1$ und $\tilde{F}_2$ sind unten zu sehen. Hierbei sind die ungeglätteten Profile schwarz und die geglätteten rot gekennzeichnet.

Um eine simultane Analyse der zwei Datensätze D_1 und D_2 zu ermöglichen, werden diese analog zu Abschnitt 4.1.3 zu einem Datensatz

$$D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \tag{7.2}$$

zusammengefügt. Wegen der identischen Wellenlängenbereiche kann dies ohne weitere Vorbehandlung der Daten durchgeführt werden. Bei der Nutzung des gegebenen photokinetischen Modells ist zu beachten, dass die numerische Auswertung des Anfangswertproblems für jeden Datensatz mit den entsprechenden Startkonzentrationen einzeln durchgeführt werden muss. Anschließend werden diese analog zu Gleichung (7.2) zusammengefügt. Bei der numerischen Auswertung spielt der photokinetische Faktor $F(t)$ eine wichtige Rolle. Eine Approximation ist unter Kenntnis der Anregungswellenlänge von 405nm möglich. Hierfür werden die zeitlichen Profile $P_1 := D_1(:, 269)$ und $P_2 := D_2(:, 269)$ zur Wellenlänge 404.98nm verwendet. Eine Approximation $\tilde{F}$ von $F(t)$ bezüglich des gegebenen Zeitgitters ist nun mittels der Formel $\tilde{F}_i = \frac{1-10^{-P_i}}{P_i}$ für $i = 1, 2$ möglich [83]. Die Ergebnisse sind in der unteren Zeile der Abbildung 7.3 als schwarze Linien zu sehen. Um den Einfluss des Rauschens zu verringern, wird zusätzlich eine Glättung von $\tilde{F}_1$ und $\tilde{F}_2$ mittels eines Savitzky-Golay Filters durchgeführt [99]. Hierzu werden Polynome vom Grad 3 genutzt. Die entsprechende Fensterbreite hängt von der Länge des jeweiligen Zeitgitters ab und beträgt 25 für D_1 und 36 für D_2. In den unteren Grafiken der Abbildung 7.3 sind die Resultate als rot gestrichelte Linie eingezeichnet. Da die numerische Aus-

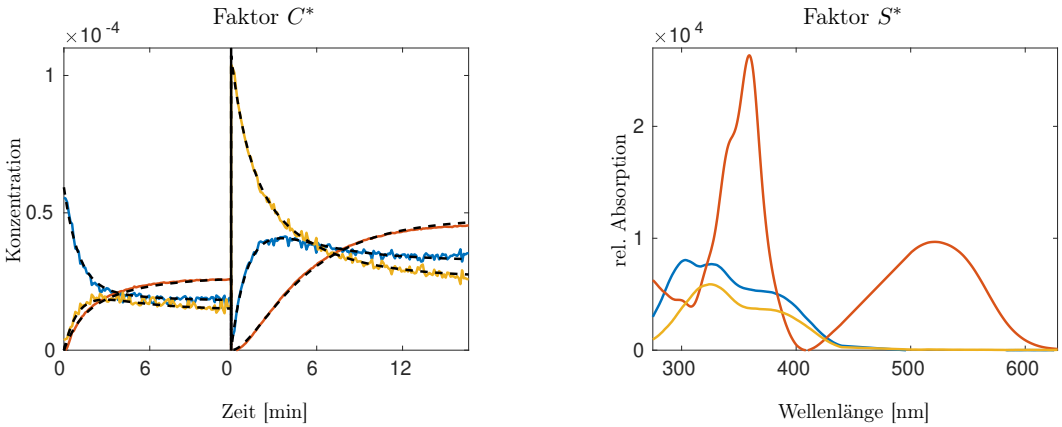


Abbildung 7.4.: Ergebnis der Matrixfaktorisierung des in Abbildung 7.3 gezeigte, zusammengesetzten Datensatzes D, siehe Gleichung (7.2). Links sind die Konzentrationsverläufe in Farbe und das mit ϕ^* parametrierte kinetische Modell als gestrichelte, schwarze Linien dargestellt. Die ermittelten Spektren sind mit der gleichen Farbcodierung rechts zu sehen.

wertung des zugrunde liegenden Anfangswertproblems die Kenntnis des photokinetischen Faktors für beliebige Zeitpunkte t voraussetzt, wird eine lineare Interpolation auf Basis der approximierten photokinetischen Faktoren $\tilde{F}_1$ und $\tilde{F}_2$ genutzt.

Zur Minimierung der Zielfunktion (3.6) aus Abschnitt 3.1 mit den Parametern $\gamma_1 = \gamma_2 = \gamma_3 = \delta_1 = 1$ wird die Matlab-Routine *lsqnonlin* mit den Parametern $TolX = TolFun = 10^{-10}$ verwendet. Die Optimierung wird mit

$$\phi^{(0)} = (\phi_1^{(0)}, \phi_{-1}^{(0)}, \phi_2^{(0)}, \phi_{-2}^{(0)}) = (0.2, 0.17, 0.07, 0.34)$$

initialisiert. Die optimierten Quantenausbeuten lauten

$$\phi^* = (\phi_1^*, \phi_{-1}^*, \phi_2^*, \phi_{-2}^*) = (0.1432, 0.2785, 0.2258, 0.3681).$$

Die Ergebnisse sind in Abbildung 7.4 zu sehen. Die Residuen der ermittelten Faktorisierung und Kinetikanpassung lauten

$$\begin{aligned} \frac{\|D - C^*(S^*)^T\|_F}{\|D\|_F} &= 0.0053, & \frac{\|C^* - C^{\mathrm{dgl}}(\phi^*)\|_F}{\|C^*\|_F} &= 0.031, \\ \frac{\|\min(C^*, 0)\|_F}{\|C^*\|_F} &= 0.00072, & \frac{\|\min(S^*, 0)\|_F}{\|S^*\|_F} &= 0.0013. \end{aligned} \tag{7.3}$$

Mittels der in Kapitel 4 vorgestellten theoretischen Ergebnisse wird nun eine Analyse mittels der Menge D-konsistenter Parameter $\mathcal{K}$ durchgeführt. Hierzu werden zwei Ansätze genutzt. Als erstes wird der Würfeleinschließungsalgorithmus aus Abschnitt 6.2 zur Bestimmungen einer Approximation $\tilde{\mathcal{K}}$ der Menge $\mathcal{K}$ verwendet. Der Algorithmus wird nicht im Raum der Quantenausbeuten angewandt, sondern es wird zunächst eine Umskalierung $k = \mathrm{diag}(\varepsilon_{\mathcal{X}}, \varepsilon_{\mathcal{Y}}, \varepsilon_{\mathcal{X}}, \varepsilon_{\mathcal{Z}})\phi$ vorgenommen. Unter Verwendung der Rückskalierung $\phi = \mathrm{diag}(\varepsilon_{\mathcal{X}}, \varepsilon_{\mathcal{Y}}, \varepsilon_{\mathcal{X}}, \varepsilon_{\mathcal{Z}})^{-1}k$ lässt sich die Menge $\mathcal{K}$ als Niveaumenge der Funktion $F_{\mathrm{eig}}(k)$ (siehe (4.14)) zum Wert 0 definieren. Für den Würfeleinschließungsalgorithmus werden die Kantenlänge $\omega = 30$ und eine Fehlertoleranz $\varepsilon = 10^{-8}$ genutzt, wobei auf weitere Verfeinerungsschritte verzichtet wird. Die Approximation $\tilde{\mathcal{K}}$ ist in der linken Grafik der Abbildung 7.5 zu sehen. Als zweites wird die in Abschnitt 4.1.2 beschriebene analytische Darstellung von $\mathcal{K}$ bestimmt. Letztere kann leicht durch Permutation der Quantenausbeuten aus den Betrachtungen zum kinetischen Modell $\mathcal{X} \rightleftharpoons \mathcal{Y} \rightleftharpoons \mathcal{Z}$ hergeleitet

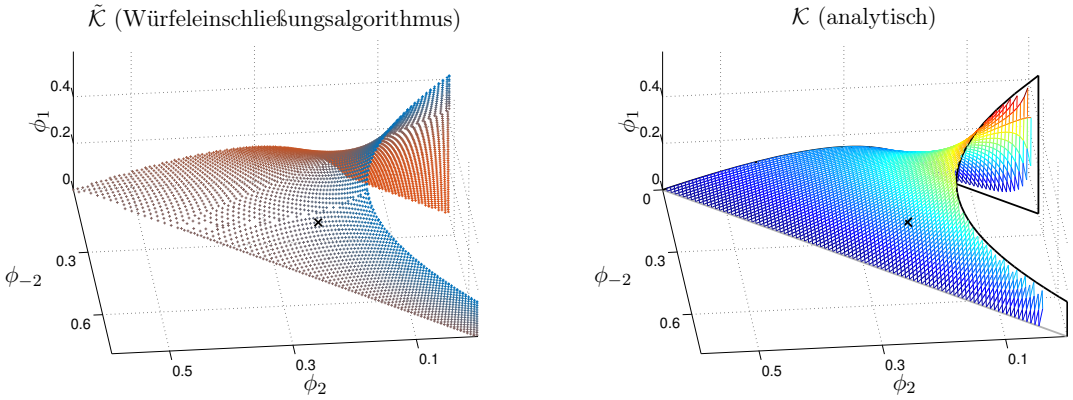


Abbildung 7.5.: Menge der D-konsistenter Parameter $\mathcal{K}$. In der linken Grafik sind die mittels Würfeleinschließungsalgorithmus berechneten Repräsentanten der Menge $\mathcal{K}$ dargestellt. Die rechte Grafik zeigt die Auswertung der analytischen Beschreibung der Menge $\mathcal{K}$ durch die Gleichung (7.4). Als Kreuz ist jeweils ϕ^* gekennzeichnet.

werden. Eine detaillierte Beschreibung ist in [111] zu finden. Seien hierzu $\lambda_1 = -1846.1$ und $\lambda_2 = -463.2$ die von 0 verschiedenen Eigenwerte von $M(\phi^*)$. Die Gleichung

$$\phi_1(\phi_2, \phi_{-2}) = -\frac{1}{\varepsilon_{\mathcal{X}}^2 \phi_2} \cdot (\varepsilon_{\mathcal{X}}\phi_2 + \varepsilon_{\mathcal{Z}}\phi_{-2} + \lambda_1) \cdot (\varepsilon_{\mathcal{X}}\phi_2 + \varepsilon_{\mathcal{Z}}\phi_{-2} + \lambda_2) \tag{7.4}$$

kann nun zur analytischen Beschreibung von $\mathcal{K}$ genutzt werden. Die Ergebnisse sind in der rechten Grafik der Abbildung 7.5 dargestellt. Die Möglichkeit für dieses kinetische Modell neben der Approximation auch eine analytische Darstellung von $\mathcal{K}$ herzuleiten, wird nun zur Verifikation der Approximation genutzt. Dafür erfolgt nun die Auswertung der Elemente aus $\tilde{\mathcal{K}}$ mittels Gleichung (7.4):

$$\max_{\tilde{\phi} \in \tilde{\mathcal{K}}} \frac{|\tilde{\phi}_1 - \phi_1(\tilde{\phi}_2, \tilde{\phi}_{-2})|}{\sum_i \tilde{\phi}_i} \approx 2.73 \cdot 10^{-4}.$$

Es sei angemerkt, dass die im Nenner stehende Summe der Quantenausbeuten anders als bei "klassischen" kinetischen Modellen nicht konstant sein muss. Der maximale relative Fehler der Elemente der Approximation $\tilde{\mathcal{K}}$ bezüglich der analytischen Lösung $\mathcal{K}$ ist offensichtlich klein, sodass von einer guten Approximation ausgegangen werden kann.

Die Menge $\tilde{\mathcal{K}}$ umfasst eine ganze Reihe Punkte, welche gar nicht auf einen nichtnegativen Faktor S führen. Durch Berechnung des Spektrenfaktors S für jedes Element von $\tilde{\mathcal{K}}$ und dessen Auswertung bezüglich Nichtnegativität wird die Reduktion auf die Menge $\tilde{\mathcal{K}}^+$ bestimmt. Da die Datensätze D_1 und D_2 durch Rauschen gestört sind, wird eine Toleranz von 2.2% komponentenweiser negativer Beiträge für den Faktor S zugelassen. Es muss also analog zu Definition 4 die Bedingung

$$\frac{S_{i,j}}{\max_l(|S_{l,j}|)} + 0.022 \geq 0 \quad \forall i, j \tag{7.5}$$

erfüllt sein. Weil Quantenausbeuten einen Wert im Intervall $[0, 1]$ annehmen müssen, kann eine weitere Einschränkung von $\tilde{\mathcal{K}}^+$ auf $\tilde{\mathcal{K}}^{+,1} = \tilde{\mathcal{K}}^+ \cap \{\tilde{\phi} \in \tilde{\mathcal{K}} : \tilde{\phi} \in [0, 1]^4\}$ durchgeführt werden. Diese Menge und der Einfluss der Verwendung von mehreren Datensätzen ist in Abbildung 7.6 dargestellt. Trotz der Einschränkung durch eine Kinetik und die

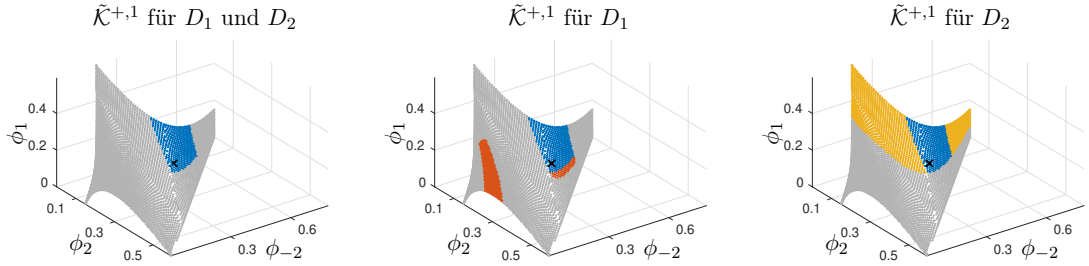


Abbildung 7.6.: Die Darstellungen zeigen als Vereinigung der farblich markierten Punkte die Mengen $\tilde{\mathcal{K}}^{+,1}$ unter Berücksichtigung der Nichtnegativität der Spektren für beide Datensätze (links), für D_1 (Mitte) und für D_2 (rechts). In grau sind jeweils die restlichen Element der Menge $\tilde{\mathcal{K}}$ gezeigt. Es gilt die Beziehung, dass der Schnitt der Mengen $\tilde{\mathcal{K}}^{+,1}$ für D_1 oder D_2 gerade der Menge $\tilde{\mathcal{K}}^{+,1}$ unter Berücksichtigung der Nichtnegativität für beide Datensätze entspricht. Als Kreuz ist ϕ^* gekennzeichnet.

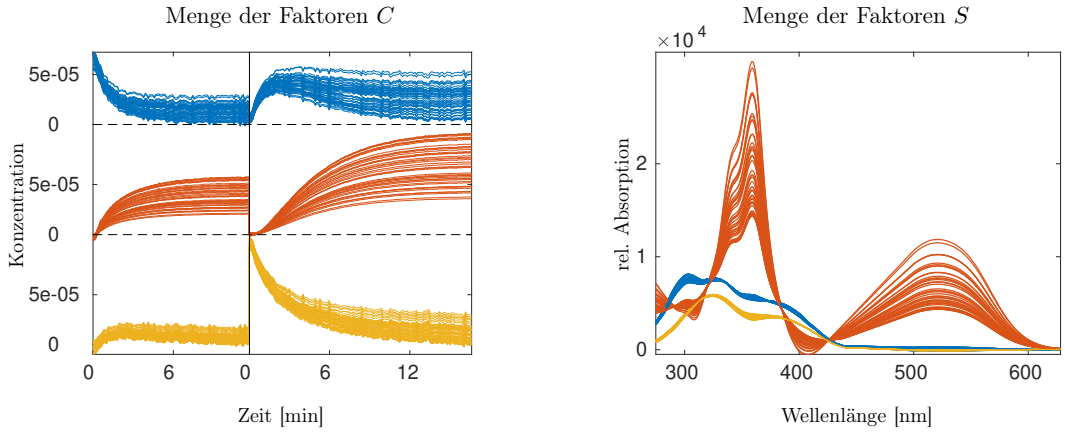


Abbildung 7.7.: Darstellung der Menge von Faktoren C (links) und S (rechts) zu einem jeden Punkt der Approximation $\tilde{\mathcal{K}}^{+,1}$ aus Abbildung 7.6.

Nichtnegativität für S sind immer noch qualitativ und quantitativ höchst unterschiedliche Zerlegungen möglich. Um dies zu verdeutlichen, sind in Abbildung 7.7 die Faktoren C und S zu allen noch in $\tilde{\mathcal{K}}^{+,1}$ verbliebenen Vektoren von Quantenausbeuten dargestellt. In Abbildung 7.8 ist die Menge der Faktoren S in niedrigdimensionaler Darstellung (schwarze Kreuze) in der Menge zulässiger Lösungen $\mathcal{M}$ gezeigt. Dies kann als Projektion der Menge $\tilde{\mathcal{K}}^{+,1}$ in $\mathcal{M}$ verstanden werden. Zur Berechnung der Menge zulässiger Lösungen $\mathcal{M}$ wurde eine Toleranz der Nichtnegativität von 3% zugelassen. Zusammen mit Abbildung 7.7 ist zu erkennen, dass die gewählte Toleranz für die Faktoren S der schwarz in $\mathcal{M}$ eingezeichneten Kreuze ausreichend ist, da diese sich innerhalb der äußeren Ränder der Segmente befinden [13, 106]. Für die Faktoren C ist der Toleranzwert nicht ausreichend. Dies ist in beiden Abbildungen an der mit rot markierten Komponente zu erkennen, da insbesondere die schwarzen Punkte in $\mathcal{M}$ die Segmentgrenzen nach innen überschreiten und somit kein komplementärer Faktor (hier C) gefunden werden kann der die Nichtnegativität unter der gegebenen Toleranz erfüllt [101].

Nun wird die Menge $\tilde{\mathcal{K}}^{+,1}$ nach Quantenausbeuten durchsucht, deren zugeordnete Faktorisierung von D kleinere Residuennormen aufweisen als C^* und S^*, siehe (7.3). Hierzu werden alle Elemente der Menge $\tilde{\mathcal{K}}^{+,1}$ bezüglich des Rekonstruktions- und Anpassungsfehlers sowie der Nichtnegativität des Konzentrationsfaktors untersucht. In Tabelle 7.1 sind die Ergebnisse für den Referenzvektor ϕ^* sowie untere und obere Schranken bezüg-

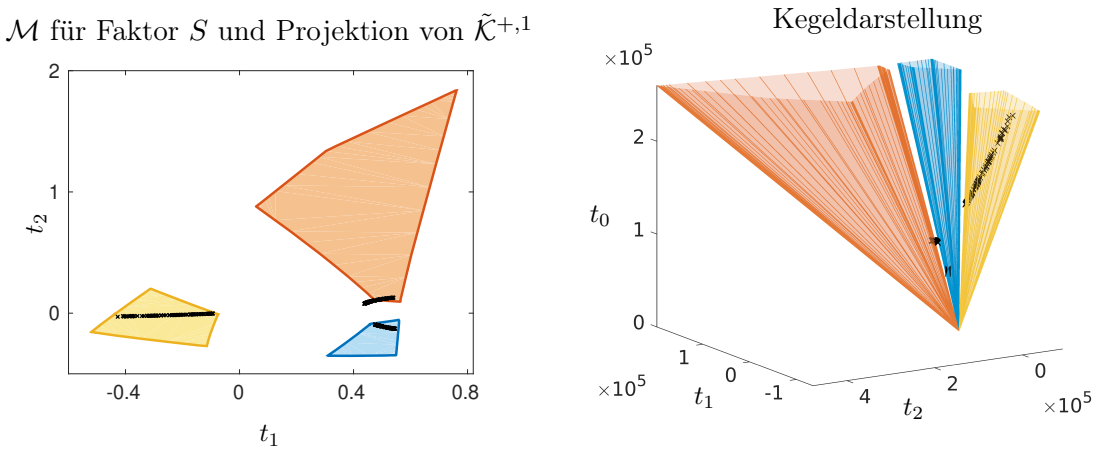


Abbildung 7.8.: Links ist die Projektion der Menge $\tilde{\mathcal{K}}^{+,1}$ (schwarz) in $\mathcal{M}$ dargestellt. Rechts ist eine Kegeldarstellung gezeigt. Die Menge $\mathcal{M}$ entspricht dem Schnitt einer (durch die Skalierung von S festgelegten) Ebene mit diesen Kegeln.

	$\frac{\|D-CS^T\|_F}{\|D\|_F}$	$\frac{\|C-C^{\mathrm{dgl}}\|_F}{\|C\|_F}$	$\frac{\|\min(C,0)\|_F}{\|C\|_F}$
Referenz ϕ^*	0.0053	0.031	$7.2 \cdot 10^{-4}$
obere Schranke $\forall\tilde{\phi} \in \tilde{\mathcal{K}}^{+,1}$	0.0053	0.072	$1.9 \cdot 10^{-2}$
untere Schranke $\forall\tilde{\phi} \in \tilde{\mathcal{K}}^{+,1}$	0.0053	0.029	$6.24 \cdot 10^{-4}$
$\phi^{\mathrm{opt}} \in \tilde{\mathcal{K}}^{+,1}$	0.0053	0.030	$6.24 \cdot 10^{-4}$

Tabelle 7.1.: Ergebnisse der Auswertung von $\tilde{\mathcal{K}}^{+,1}$ bezüglich verschiedener Fehlerindikatoren.

lich der gesamten Menge $\tilde{\mathcal{K}}^{+,1}$ angegeben. Darüber hinaus kann mit

$$\phi^{\mathrm{opt}} = (0.1372, 0.2898, 0.2318, 0.3629)$$

ein Vektor von Quantenausbeuten bestimmt werden, der bezüglich aller Indikatoren einen geringeren Fehlerwert aufweist als ϕ^*. Die entsprechende Zerlegung ist in Abbildung 7.9 dargestellt.

Abschließend wird die Menge $\mathcal{K}^+_{\varepsilon,\theta}$ analysiert. Es wird nun der Fehler der Kinetikanpassung auf Basis von ϕ^* für die Menge $\mathcal{K}(\phi^*)$ nach Abschnitt 5.1 abgeschätzt. Mit den Folgerungen 1 und 2 ergibt sich eine obere Schranke des absoluten Fehlers

$$\|C(\phi)-C^{\mathrm{dgl}}(\phi)\|_F \leq \|C(\phi^*)-C^{\mathrm{dgl}}(\phi^*)\|_F \|T_{\phi^*} T_k^+\|_F = 0.425 \cdot 3.14 = 1.3345 \;\; \forall \phi \in \mathcal{K}(\phi^*).$$

Sie kann als Indikator für den relativen Fehler genutzt werden, womit in etwa mit einer Verdreifachung des Fehlers zu rechnen ist. Weil der relative Fehler der Kinetikanpassung bezüglich ϕ^* nach Tabelle 7.1 mit 3% bereits hoch ist, führt $\mathcal{K}(\phi^*)$ nur lokal um ϕ^* auf Geschwindigkeitsparameter mit ähnlich guten oder sogar besseren Fehlerindikatoren (siehe ϕ^{opt}). Es wird daher empfohlen eine detailliertere Analyse durch Betrachten der Menge zulässiger D-approximativer Parameter durchzuführen.

Hierzu wird der Würfeleinschließungsalgorithmus eingesetzt um eine Approximation $\tilde{\mathcal{K}}^+_{\varepsilon,\theta}$ von $\mathcal{K}^+_{\varepsilon,\theta}$ zu bestimmen. Es werden die Niveaumengen der Funktion $F_{\varepsilon,\theta}(k)$ zu $\theta = 0.022$ und $\varepsilon \in \{0.012, 0.014, 0.016, 0.018\}$ berechnet (siehe (5.2)). Der Parameter k ergibt sich erneut durch die Umskalierung $k = \mathrm{diag}(\varepsilon_{\mathcal{X}}, \varepsilon_{\mathcal{Y}}, \varepsilon_{\mathcal{X}}, \varepsilon_{\mathcal{Z}})\phi$ mittels der molaren Absorptionskoeffizienten. Die Gewichtung der Zielfunktion lautet $\gamma_1 = \gamma_2 = \delta_1 = 1$ und für den

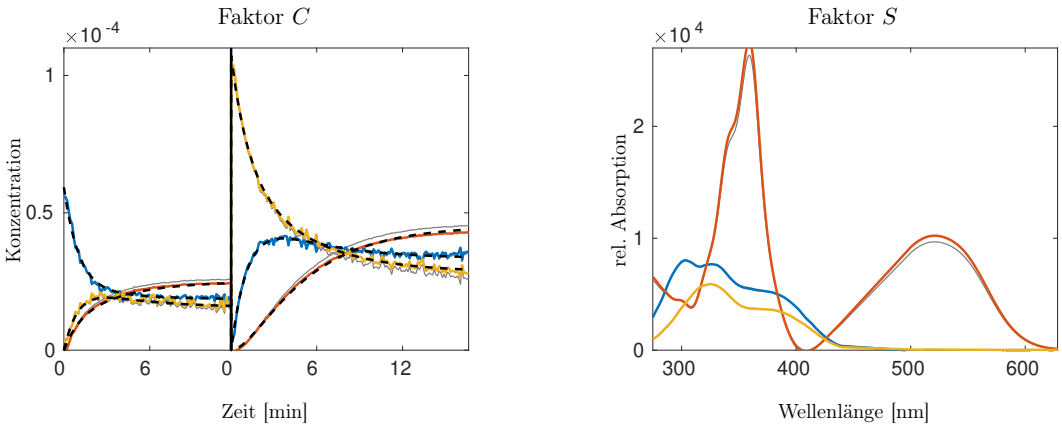


Abbildung 7.9.: Vergleich der Zerlegungen, welche den Quantenausbeuten ϕ^{opt} (farbig) und ϕ^* (grau) zugeordnet werden können. Mit ϕ^{opt} konnte gegenüber ϕ^* eine Verbesserung bezüglich aller betrachteter Fehler erzielt werden, siehe Tabelle 7.1.

Würfeleinschließungsalgorithmus werden die Kantenlänge $\omega = 40$ und $\varepsilon = 10^{-5}$ genutzt, wobei auf weitere Verfeinerungsschritte verzichtet wird. Weil eine Darstellung der resultierenden Elemente $\phi \in \mathcal{K}^+_{\varepsilon,\theta}$ (nach erneuter Umskalierung) ohne die zuvor verwendete Dimensionsreduktion im vierdimensionalen nicht möglich ist, werden diese durch Weglassen von ϕ_{-1} in den bereits in Abbildung 7.5 gezeigten Raum projiziert. Zusätzlich werden nicht die Repräsentanten selbst sondern zur Verbesserung der Übersicht deren konvexe Hülle gezeigt. Die vier berechneten Mengen $\mathcal{K}^+_{\varepsilon,\theta}$ sind in Abbildung 7.10 dargestellt. Ein Expandieren der Menge mit größer werdender Fehlertoleranz ε ist gut zu erkennt. Zur besseren räumlichen Vorstellung ist die Menge $\tilde{\mathcal{K}}^+_{0.014,0.022}$ nochmals aus verschiedenen Blickwinkeln in Abbildung A.3 des Anhangs A.3 dargestellt. Analog zur Abbildung 7.7 kann auch die Menge $\mathcal{K}^+_{\varepsilon,\theta}$ in den Raum möglicher Faktoren C und S übertragen werden, siehe Abbildung A.4 in Anhang A.3.

7.3.2. Ligandensubstitutionsreaktion

Im Folgenden wird eine Ligandensubstitutionsreaktion von Iridium-Hydridokomplexen im Rahmen von in situ spektroskopischen Untersuchungen zu Hydroformylierungsreaktionen untersucht. Die Messserie wurde durch Dr. Christoph Kubis des LIKAT in Rostock bereitgestellt. Es werden nun sowohl eine regularisierte Matrixfaktorisierung mit simultaner Anpassung der Kinetikparametrierungen als auch Approximationen der daraus resultierenden Mengen D-konsistenter Parameter $\mathcal{K}$, zulässiger Parameter $\mathcal{K}^+$ und zulässiger D-approximativer Parameter $\mathcal{K}^+_{\varepsilon,\theta}$ bestimmt.

Spektroskopischer Datensatz 3. Es handelt sich um ein reversibles Reaktionssystem bei dem die chemische Komponente $\mathcal{X}$=$\text{HIr(CO)}_3\text{L}$ unter Zugabe des Liganden L zur Komponente $\mathcal{Y}$=$\text{HIr(CO)}_2\text{L}_2$ reagiert, wobei sich schlussendlich ein Gleichgewichtszustand zwischen $\mathcal{X}$ und $\mathcal{Y}$ einstellt. Zur Vorbereitung der weiteren Analyse der Spektrenfolge waren nur wenige Datenvorbehandlungsschritte nötig. Neben der Subtraktion eines Lösungsmittelspektrums von jedem Spektrum der Messfolge war lediglich eine geringfügige Grundlinienkorrektur durchzuführen. Die resultierende Matrix $D \in \mathbb{R}^{m\times n}$ der Messserie enthält $m = 1064$ Spektren in einem Zeitintervall $[1.33\text{min},\ 1118.0\text{min}]$ zu jeweils $n = 1009$ Wellenzahlen im Intervall $[1957\text{cm}^{-1},\ 2200\text{cm}^{-1}]$. Die vorbehandelte Matrix D ist in Abbildung 7.11 dargestellt. Die weiten Systemeigenschaften lauten:

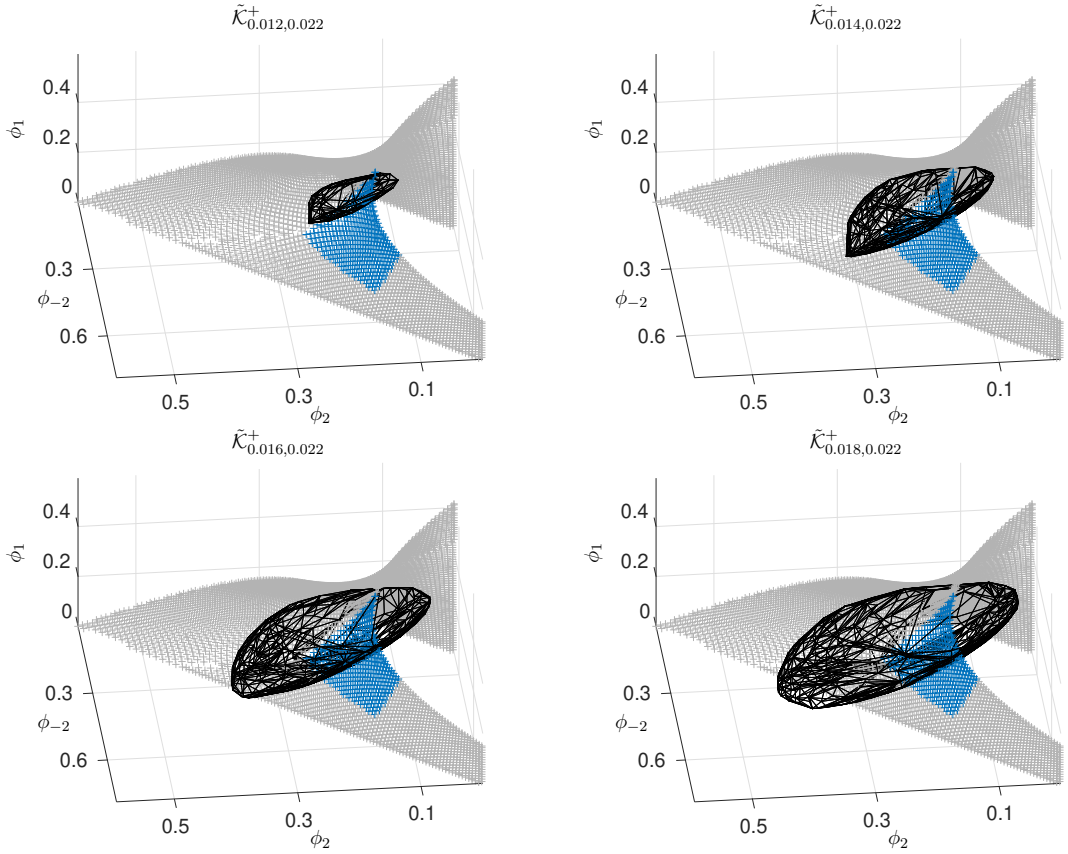


Abbildung 7.10.: Es sind die konvexen Hüllen der vier Mengen $\mathcal{K}^+_{\varepsilon,\theta}$ zu den jeweiligen Fehlertoleranzen in schwarz dargestellt. In grau ist $\mathcal{K} \setminus \mathcal{K}^+$ und in blau $\mathcal{K}^+$ zu sehen. Größere Fehlertoleranzen führen erwartungsgemäß zu einer Ausdehnung der Mengen.

- Gesamtdruck $P(\mathrm{CO}/\mathrm{H}_2) = 1$ bar, Iridiumkonzentration [Ir]=20 mmol/ℓ mit [L]/[Ir]= 4, Lösungsmittel: Toluol.
- Kinetisches Modell: $\mathcal{X} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y}$ mit den zugeordneten Konzentrationsverläufe $x(t)$ und $y(t)$.
- Zur vereinfachten Analyse werden die Startkonzentrationen $c_0 = (1, 0)$ angenommen. Die Verläufe $x(t)$ und $y(t)$ beschreiben also den Anteil von $\mathcal{X}$ und $\mathcal{Y}$ am Gesamtgemisch. Damit kann weiter auf die Angabe der Einheiten der Konzentrationen und der Geschwindigkeitsparameter $k \in \mathbb{R}^2$ verzichtet werden.
- Differentialgleichungssystem:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} -k_1 & k_{-1} \\ k_1 & -k_{-1} \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$$

Zur Minimierung der Zielfunktion (3.6) aus Abschnitt 3.1 mit $\gamma_1 = \gamma_2 = \gamma_3 = \delta_1 = 1$ wird die Matlab-Routine *lsqnonlin* mit den Fehlertoleranzen $TolX = TolFun = 10^{-10}$ verwendet. Die Optimierung wird mit

$$k^{(0)} = (0.01, 0.01)^T$$

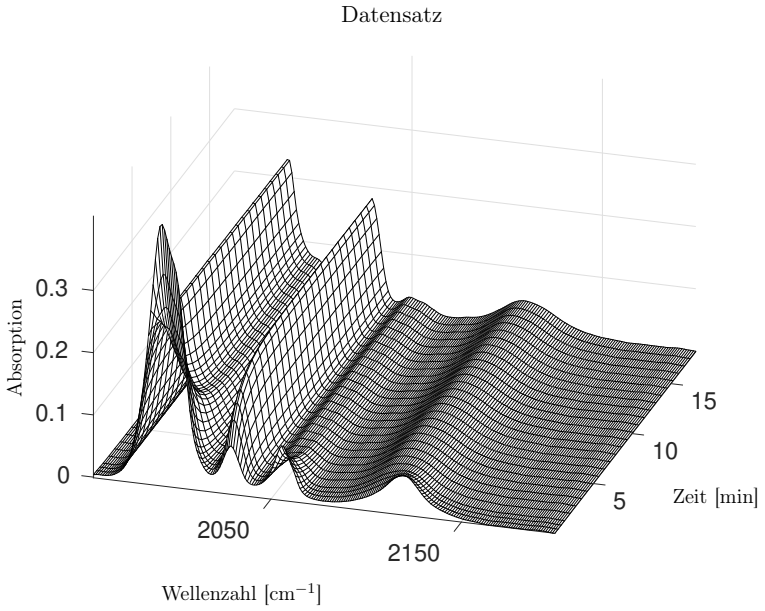


Abbildung 7.11.: Darstellung der FTIR-Messserie des spektroskopischen Datensatzes 3 zu einer Ligandensubstitutionsreaktion.

initialisiert und die optimierten Geschwindigkeitsparameter lauten

$$k^* = (k_1^*, k_{-1}^*)^T = (0.01038, 0.00277)^T \ .$$

Die resultierenden Faktoren sind in Abbildung 7.12 gezeigt. Die Residuen zur ermittelten Faktorisierung und Kinetikanpassung lauten

$$\frac{\|D - C^*(S^*)^T\|_F}{\|D\|_F} = 0.0036, \qquad \frac{\|C^* - C^{\mathrm{dgl}}(k^*)\|_F}{\|C^*\|_F} = 0.0092,$$
$$\frac{\|\min(C^*, 0)\|_F}{\|C^*\|_F} = 0.00042, \qquad \frac{\|\min(S^*, 0)\|_F}{\|S^*\|_F} = 0.00077,$$

womit von einer guten Approximation ausgegangen wird. Nun wird der Würfeleinschließungsalgorithmus zur Bestimmungen einer Approximation der Menge $\mathcal{K}$ unter der Annahme $k^* \in \mathcal{K}$ verwendet. Analog zum vorherigen Beispiel erfolgt die Beschreibung von $\mathcal{K}$ als Niveaumenge der Funktion $F_{\mathrm{eig}}(k)$ nach (4.14) zum Wert 0. Für den Würfeleinschließungsalgorithmus werden die Kantenlänge $\omega = (k_1^* + k_{-1}^*)/100 \approx 1.315 \cdot 10^{-4}$ und eine Fehlertoleranz $\varepsilon = 10^{-6}$ genutzt, wobei auf weitere Verfeinerungsschritte verzichtet wird. Die Approximation $\tilde{\mathcal{K}}$ von $\mathcal{K}$ ist in der linken Grafik der Abbildung 7.13 als Vereinigung der schwarzen und grünen Punkte dargestellt. Darüber hinaus ist in der selben Grafik eine Approximation der Menge $\mathcal{K}^+$ (grüne Punkte) dargestellt. Diese wird durch Auswertung eines jeden Repräsentanten von $\mathcal{K}$ bezüglich der Negativität des jeweiligen Faktors S bestimmt. Es wird analog zu (7.5) die Bedingung

$$\frac{S_{i,j}}{\max_l(|S_{l,j}|)} + 0.011 \geq 0 \quad \forall i, j$$

verwendet. Dies entspricht einer komponentenweiser Toleranz negativer Einträge in S von 1.1%.

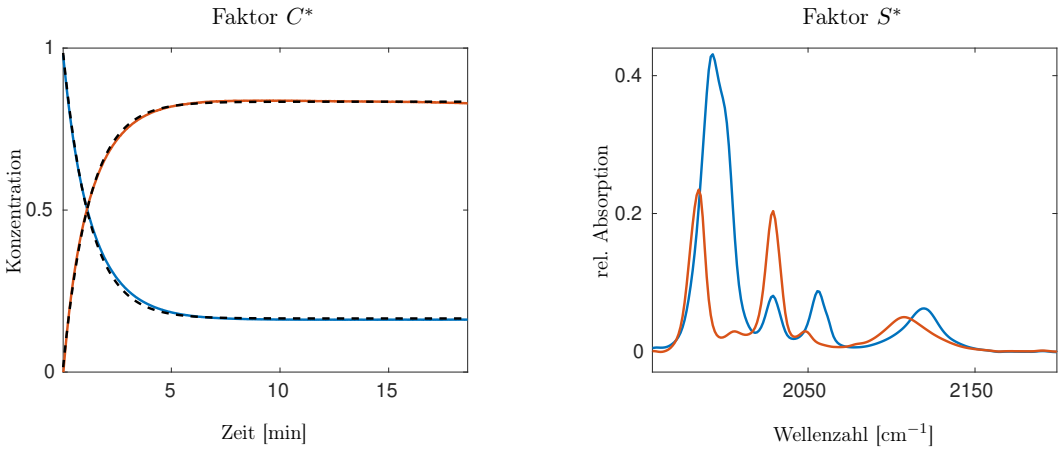


Abbildung 7.12.: Ergebnis der Zerlegung des spektroskopischen Datensatzes 3. Links sind die Konzentrationsverläufe in Farbe und das mit k^* parametrierte kinetische Modell als gestrichelte, schwarze Linien dargestellt. Die ermittelten Spektren sind mit der gleichen Farbcodierung rechts dargestellt.

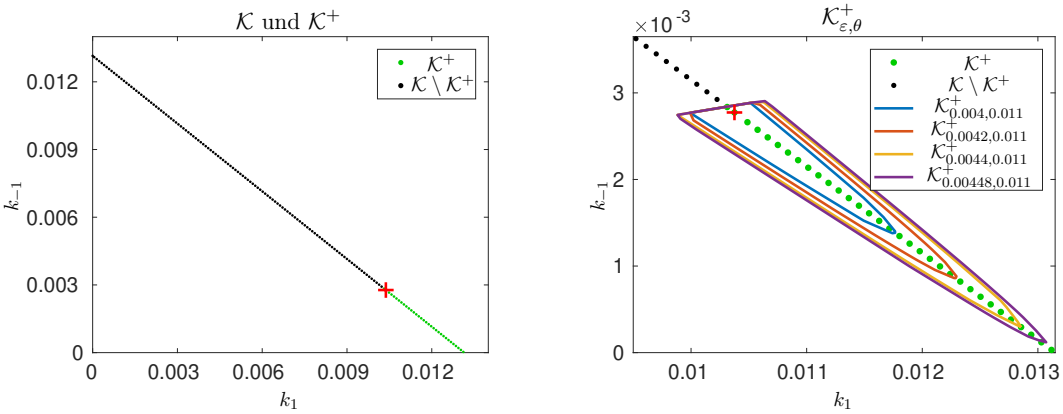


Abbildung 7.13.: Repräsentationen der Mengen $\mathcal{K}$ und $\mathcal{K}^+$ (links) sowie der Menge $\mathcal{K}^+_{\varepsilon,\theta}$ (rechts) zu $\theta = 0.011$ und $\varepsilon \in \{0.004, 0.0042, 0.0044, 0.00448\}$. In der rechten Grafik ist die konvexe Hülle der durch den Würfeleinschließungsalgorithmus bestimmen Repräsentanten der jeweiligen Menge $\mathcal{K}^+_{\varepsilon,\theta}$ gezeigt. Es ist leicht zu erkennen, dass steigenden Fehlertoleranzen ε zu einem Expandieren der Menge $\mathcal{K}^+_{\varepsilon,\theta}$ führen. Der Geschwindigkeitsparameter k^* ist als rotes Kreuz in beiden Grafiken eingezeichnet.

Analog zum vorherigen Beispiel wird abschließend die Menge $\mathcal{K}^+_{\varepsilon,\theta}$ analysiert. Es wird der Fehler der Kinetikanpassung auf Basis von k^* für $\mathcal{K}(k^*)$ nach Abschnitt 5.1 abgeschätzt. Aus den Folgerungen 1 und 2 resultiert eine obere Schranke des absoluten Fehlers

$$\|C(k) - C^{\mathrm{dgl}}(k)\|_F \leq \|C(k^*) - C^{\mathrm{dgl}}(k^*)\|_F \|T_{k^*} T_k^+\|_F = 0.0076 \cdot 2.72 = 0.0207 \ \ \forall k \in \mathcal{K}(k^*).$$

Sie kann als Indikator für den relativen Fehler genutzt werden, womit erneut in etwa mit einer Verdreifachung des Fehlers zu rechnen ist. Weil der relative Fehler der Kinetikanpassung bezüglich k^* im Gegensatz zum vorherigen Beispiel nur 0.76% beträgt, kann mit $\mathcal{K}(k^*)$ sowie der daraus resultierenden Menge $\mathcal{K}^+$ von einer guten Approximation für $\mathcal{K}^+_{\varepsilon,\theta}$ ausgegangen werden. Um dies zu bestätigen folgt nun eine detailliertere Analyse der Menge zulässiger D-approximativer Parameter.

Der Würfeleinschließungsalgorithmus wird eingesetzt um eine Approximation $\tilde{\mathcal{K}}^+_{\varepsilon,\theta}$ von $\mathcal{K}^+_{\varepsilon,\theta}$ zu bestimmen. Es werden die Niveaumengen der Funktion $F_{\varepsilon,\theta}(k)$ zu $\theta = 0.011$ und $\varepsilon \in \{0.004, 0.0042, 0.0044, 0.00448\}$ berechnet (siehe (5.2)). Die Gewichte der Zielfunktion lauten $\gamma_1 = \gamma_2 = \delta_1 = 1$. Für den Würfeleinschließungsalgorithmus wird die

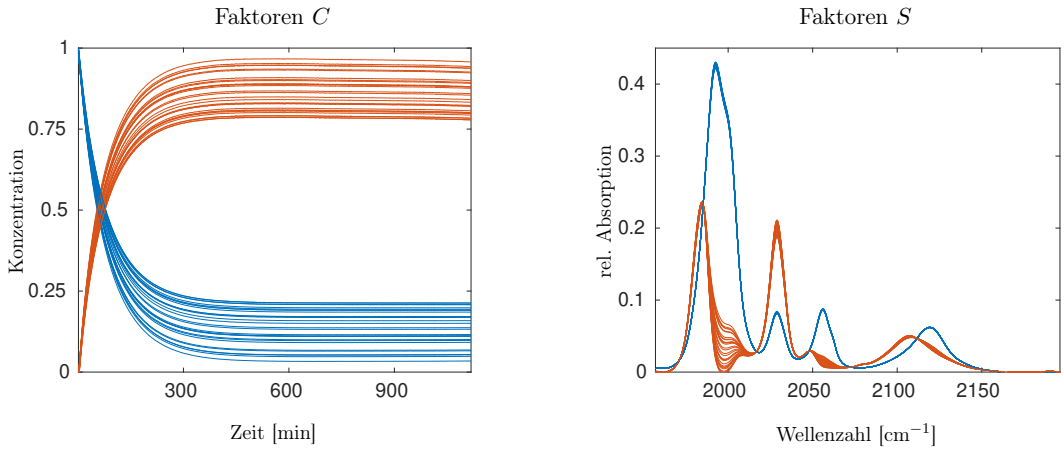


Abbildung 7.14.: Es sind die Faktoren C und S für 20% der durch den Würfeleinschließungsalgorithmus bestimmen, Repräsentanten von $\mathcal{K}^+_{0.00448,0.011}$ dargestellt. Die möglichen Faktoren C weisen untereinander eine starke Variation auf. Im Gegensatz dazu ist die erste Spalte des Faktors S (in blau) nahezu eindeutig bestimmt. Die zweiten Spalten der möglichen Faktoren S (in rot) zeigen in lokal begrenzte Wellenzahlintervallen Variationen auf.

Kantenlänge $\omega = 1.315 \cdot 10^{-4}$ genutzt und auf weitere Verfeinerungsschritte verzichtet. Die Approximationen der Mengen $\tilde{\mathcal{K}}^+_{\varepsilon,\theta}$ sind in der rechten Grafik der Abbildung 7.13 dargestellt. Sie ähneln stark der durch grüne Punkte angedeuteten Menge $\mathcal{K}^+$ ohne das ein starkes Expandieren in dazu orthogonaler Richtung stattfindet, wie im vorherigen Anwendungsbeispiel. Diese Beobachtung unterstützt die Aussage, dass es sich bei $\mathcal{K}^+$ um eine gute Approximation an $\mathcal{K}^+_{\varepsilon,\theta}$ handelt. Um eine bessere visuelle Auswertung von beispielsweise $\mathcal{K}^+_{0.00448,0.011}$ zu ermöglichen, sind zusätzlich in Abbildung 7.14 die Bänder der entsprechenden Faktoren C und S gezeigt.

7.4. Kritische Zusammenfassung

In diesem Kapitel wird gezeigt, dass sich die analytischen Betrachtungen mittels der Mengen D-konsistenter und zulässiger Parameter sowie deren Verallgemeinerungen für gestörte Matrizen D auf experimentell vermessene Reaktionssysteme anwenden lassen. Mittels Würfeleinschließungsalgorithmus können Approximationen der Parameterlösungsmengen $\mathcal{K}, \mathcal{K}^+, \mathcal{K}_\varepsilon$ und $\mathcal{K}^+_{\varepsilon,\theta}$ bestimmt werden. Es verbleiben Schwierigkeiten bei der Darstellung dieser Mengen, insbesondere wenn komplexere Systeme, als das in Abschnitt 7.3.1 vorgestellte, betrachtet werden. Projektionen der Parameterlösungsmengen in den Raum der links- und rechtsseitigen Singulärvektoren beziehungsweise in den Raum der Konzentrations- und Spektrenfaktoren können dann durchgeführt werden. Solche Projektionen gehen typischerweise mit einem Verlust an Informationen einher und sollten wenn möglich vermieden werden.

8. Ausblick

Kinetische Modelle können als Nebenbedingung bei der nichtnegativen Vollrangfaktorisierung eingesetzt werden, wodurch im Regelfall eine erhebliche Reduktion der Lösungsmenge erzielt wird. Dennoch können für das regularisierte Problem nichttriviale Lösungsuneindeutigkeiten existieren. Die entsprechenden Lösungsmengen lassen sich niedrigdimensional im Raum der Parameter des jeweiligen kinetischen Modells darstellen und analysieren. Die hierfür eingeführte Menge D-konsistenter Parameter $\mathcal{K}$ ist Ausgangspunkt weiterer Betrachtungen.

In dieser Arbeit werden vorrangig kinetische Modelle erster Ordnung, der Einfluss von Störungen der Matrix D auf $\mathcal{K}$ und die effiziente numerische Approximation solcher Parameterlösungsmengen untersucht. Daran anknüpfende Forschungsthemen werden im Folgenden kurz vorgestellt.

Parameterlösungsmengen von Kinetiken erster und zweiter Ordnung

In Kapitel 2 wird gezeigt, dass der Faktor C durch Linearkombinationen der linken Singulärvektoren von D gebildet werden kann. Auch in kinetischen Modellen, denen ein gewöhnliches Differentialgleichungssystem erster Ordnung zugrunde liegt, können ähnliche lineare Strukturen gefunden werden. Die Lösung eines entsprechenden Anfangswertproblems wird ganz analog durch geeignete Linearkombinationen der Funktionen in einem Fundamentalsystem bestimmt. Dieser Zusammenhang ist Grundlage für große Teile der in Kapitel 4 vorgestellte Theorie und die daraus resultierende effiziente Berechnung der Menge $\mathcal{K}$ für diese Klasse kinetischer Modelle. In den Anfangswertproblemen von Kinetiken zweiter Ordnung ist keine solche lineare Lösungsstruktur zu finden, sofern die Lösungen analytisch zugänglich sind. Für diesen Fall wird vermutet, dass unter Betrachtung idealisierter Ausgangsdaten D und unter Vernachlässigung trivialer Fälle eine eindeutige Faktorisierung $D = CS^T$ möglich ist. Für die Kinetik $2\mathcal{X} \to \mathcal{Y}$ ist diese Behauptung in Abschnitt 4.2.2 bewiesen. Ein Beweis für allgemeine Kinetiken zweiter Ordnung oder ein Gegenbeispiel stehen noch aus.

Kinetiken zweiter Ordnung für gestörte Ausgangsdaten D

Bereits kleine Störungen von D können dazu führen, dass nichttriviale Parameterlösungsmengen $\mathcal{K}_\varepsilon$ für kinetische Modelle zweiter Ordnung auftreten. Ein entsprechendes Beispiel ist in Abschnitt 5.2 für die Michaelis-Menten Kinetik zu finden. Für diesen Fall lässt sich die Lösungsuneindeutigkeit auf eine unzureichende zeitliche Auflösung der Spektrenfolge zurückführen. Von besonderem Interesse sind dann etwa hinreichende Kriterien, wie die Angabe einer maximalen Gitterschrittweite entlang der Zeitachse, sodass eine eindeutige Lösung der Faktorisierungsaufgabe bestimmt werden kann.

Ein Graph der Matrix $M(k)$ und die Menge $\mathcal{K}$

Die Reaktionsgleichung eines Systems aus s Komponenten kann als gerichteter Graph mit s Knoten, dessen Kanten durch die Geschwindigkeitsparameter gewichtet sind, betrachtet werden. Jede Kante ist also einer Teilreaktion zugeordnet. Die zugehörige Adjazenzmatrix ist eng verknüpft mit der Koeffizientenmatrix $M(k)$ des Anfangswertproblems. Für kinetische Modelle erster Ordnung wird ein Zusammenhang zwischen der Struktur von $M(k)$ und der Menge $\mathcal{K}$ auf Basis der Graphentheorie vermutet. Exemplarisch werden zwei Klassen von Kinetiken vorgestellt:

1. Konsekutivreaktionen besitzen eine Reaktionsgleichung der Form

$$\mathcal{X}_1 \xrightarrow{k_1} \dots \xrightarrow{k_{s-1}} \mathcal{X}_s\,.$$

   Die zugehörige Matrix $M(k)$ ist nur auf der Haupt- und ersten unteren Nebendiagonalen besetzt. Die Menge $\mathcal{K}$ einer entsprechenden Kinetik besteht immer aus bis zu $(s-1)!$ Punkten im $\mathbb{R}^s$.

2. Unter Berücksichtigung einer zusätzlichen Rückreaktion zwischen den letzten beiden Komponenten

$$\mathcal{X}_1 \xrightarrow{k_1} \dots \xrightarrow{k_2} \mathcal{X}_{s-1} \underset{k_s}{\overset{k_{s-1}}{\rightleftharpoons}} \mathcal{X}_s, \quad s \geq 2,$$

   setzt sich die entsprechende Menge $\mathcal{K}$ immer aus bis zu $(s-1)!$ eindimensionalen affinen Unterräumen des $\mathbb{R}^s$ zusammen. Die Matrix $M(k)$ ist analog zu Punkt 1 nur auf der Haupt- und ersten unteren Nebendiagonalen sowie in dem Element $(M(k))_{s-1,s}$ besetzt.

Die Approximation von $\mathcal{K}_\varepsilon$ durch $\mathcal{K}$

Der in Abschnitt 5.1 eingeführte globale Anpassungsfehler

$$\delta_{\max} = \max_{k \in \mathcal{K}(k^*)} \left\| (C(k) - C^{\mathrm{dgl}}(k)) B^{-1} \right\|_F$$

mit regulärer, nichtnegativer Skalierungsmatrix B ist wichtig für die Einschätzung der Aussagekraft der Menge $\mathcal{K} = \mathcal{K}(k^*)$ als Approximation für die Menge D-approximativer Parameter $\mathcal{K}_\varepsilon$. Die in Folgerung 2 vorgestellte Abschätzung von $\|B(T_{k^*} T_k^+) B^{-1}\|_F$ (und damit auch $\delta_{\max}$) erfolgt unabhängig von der zugrunde liegenden Kinetik. Wird T_k wie in Abschnitt 3.1 mit $T_k = ((U\Sigma)^+ C^{\mathrm{dgl}}(k))^+$ an die Kinetik gekoppelt, ist eine Verbesserung der Abschätzung zu erwarten.

Literaturverzeichnis

[1] N. Alcock, D. Benton, P. Moore. Kinetics of series first-order reactions. *Trans. Faraday Soc.*, 66:2210–2213, 1970.

[2] M. Alier, R. Tauler. Multivariate curve resolution of incomplete data multisets. *Chemom. Intell. Lab. Syst.*, 127:17–28, 2013.

[3] M. Alonso, J. González-Hernández, S. del Dedo. Application the computational method KINMODEL(AGDC) to the simultaneous determination of kinetic and analytical parameters. *Appl. Math. Comput.*, 219(12):7089–7101, 2013.

[4] F. Alsmeyer, H.-J. Koß, W. Marquardt. Indirect spectral hard modeling for the analysis of reactive and interacting mixtures. *Appl. Spectrosc.*, 58(8):975–985, 2004.

[5] P. Atkins, J. Paula. *Physikalische Chemie*. John Wiley & Sons, 2013.

[6] M. Aubury, W. Luk. Binomial filters. *J. Signal Process Sys.*, 12(1):35–50, 1996.

[7] T. Azzouz, R. Tauler. Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples. *Talanta*, 74(5):1201–1210, 2008.

[8] R. Bapat, T. Raghavan. *Nonnegative matrices and applications*, Bd. 64. Cambridge University Press, 1997.

[9] A. Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Ann. Phys.*, 162(5):78–88, 1852.

[10] M. Berry, M. Browne. Email surveillance using non-negative matrix factorization. *Comput. Math. Organ. Theory*, 11(3):249–264, 2005.

[11] M. Berry, M. Browne, A. Langville, V. Pauca, R. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data. Anal.*, 52:155–173, 2007.

[12] A. Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[13] O. Borgen, B. Kowalski. An extension of the multivariate component-resolution method to three components. *Anal. Chim. Acta*, 174(0):1–26, 1985.

[14] D. Cai, X. He, J. Han, T. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1548–1560, 2010.

[15] D. Calvetti, B. Lewis, L. Reichel. GMRES, L-curves, and discrete ill-posed problems. *BIT Numer. Math.*, 42(1):44–65, Mar 2002.

[16] D. Calvetti, S. Morigi, L. Reichel, F. Sgallari. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *J. Comput. Appl. Math.*, 123(1-2):423–446, 2000.

[17] Z. Chen, A. Cichocki. Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints. *Laboratory for Advanced Brain Signal Processing, RIKEN, Tech. Rep.*, 2005.

[18] W. Chew, E. Widjaja, M. Garland. Band-target entropy minimization (BTEM): An advanced method for recovering unknown pure component spectra. Application to the FT-IR spectra of unstable organometallic mixtures. *Organometallics*, 21(9):1982–1990, 2002.

[19] J. Chrastil. Determination of the first-order consecutive reversible reaction kinetics. *Comput. Chem.*, 17(1):103–106, 1993.

[20] A. Cichocki, R. Zdunek, A. Phan, S. Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[21] J. Cohen, U. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra Appl.*, 190:149–168, 1993.

[22] T. Coleman, Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.*, 6(2):418–445, 1996.

[23] R. de Fréin, K. Drakakis, S. Rickard, A. Cichocki. Analysis of financial data using non-negative matrix factorization. *Int. Math. Forum*, 3(38):1853–1870, 2008.

[24] A. de Juan, E. Casassas, R. Tauler. Soft modeling of analytical data. *Encyclopedia of analytical chemistry: Applications, theory and instrumentation*, 2006.

[25] A. de Juan, J. Jaumot, R. Tauler. Multivariate Curve Resolution (MCR). Solving the mixture analysis

problem. *Anal. Methods*, 6(14):4964–4976, 2014.

[26] A. de Juan, M. Maeder, M. Martínez, T. R. Combining hard- and soft-modelling to solve kinetic problems. *Chemom. Intell. Lab. Syst.*, 54:123–141, 2000.

[27] M. de Luca, G. Ragno, G. Ioele, R. Tauler. Multivariate curve resolution of incomplete fused multiset data from chromatographic and spectrophotometric analyses for drug photostability studies. *Anal. Chim. Acta*, 837:31–37, 2014.

[28] M. Dellnitz, G. Froyland, O. Junge. The algorithms behind GAIO-set oriented numerical methods for dynamical systems. *Ergodic theory, analysis, and efficient simulation of dynamical systems*, 560:145–174, 2001.

[29] J. Dennis, D. Gay, R. Welsch. Algorithm 573: An adaptive nonlinear least-squares algorithm. *ACM Trans. Math. Softw.*, 7:369–383, 1981.

[30] O. Devos, S. Aloïse, M. Sliwa, R. Métivier, J.-P. Placial, C. Ruckebusch. Multivariate curve resolution of (ultra)fast photoinduced process spectroscopy data. In C. Ruckebusch, Herausgeber, *Resolving spectral Mixtures with applications from Time-resolved spectroscopy to super-resolution imaging*, Bd. 30 von *Data Handling in Science and Technology*, S. 353 – 379. Elsevier, 2016.

[31] O. Devos, H. Schröder, M. Sliwa, J. Placial, K. Neymeyr, R. Métivier, C. Ruckebusch. Photochemical multivariate curve resolution models for the investigation of photochromic systems under continuous irradiation. *Anal. Chim. Acta*, 1053:32–42, 2019.

[32] J. Dormand, P. Prince. A family of embedded runge-kutta formulae. *J. Comput. Appl. Math.*, 6(1):19–26, 1980.

[33] R. Eberhart, Y. Shi. Comparing inertia weights and constriction factors in particle swarm optimization. *Proc. Congr. Evol. Comput.*, 1:84–88, 2000.

[34] C. Eckard, G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[35] P. Eilers, H. Boelens. Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Report*, 1(1):5, 2005.

[36] H. Engl, M. Hanke, A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, 1996.

[37] E. Fehlberg. Klassische Runge-Kutta-Formeln vierter und niedrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungsprobleme. *Computing*, 6(1-2):61–71, 1970.

[38] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. de Juan, A. Gorzsás. Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS). *Nat. Protoc.*, 10(2):217, 2015.

[39] O. Forster. *Analysis 1*. Springer, 2006.

[40] G. Frobenius. Über Matrizen aus nicht negativen Elementen. *Sitzungsber. Preuß. Akad. Wiss.*, S. 456–477, 1912.

[41] H. Gampp, M. Maeder, C. Meyer, A. Zuberbuehler. Quantification of a known component in an unknown mixture. *Anal. Chim. Acta*, 193:287–293, 1987.

[42] M. Garrido, F. Rius, M. Larrechi. Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Anal. Bioanal. Chem.*, 390(8):2059–2066, 2008.

[43] P. Gemperline, E. Cash. Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with penalty functions. *Anal. Chem.*, 75(16):4236–4243, 2003.

[44] K. Godfrey, J. DiStefano III. Identifiability of model parameter. *IFAC Proc. Vol.*, 18(5):89–114, 1985.

[45] S. Golotvin, A. Williams. Improved baseline recognition and modeling of FT NMR spectra. *J. Magn. Reson.*, (1):122–125, 2000.

[46] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, R. Tauler. A review of recent methods for the determination of ranges of feasible solutions from soft modeling analyses of multivariate data. *Anal. Chim. Acta*, 911:1–13, 2016.

[47] A. Golshan, H. Abdollahi, M. Maeder. Resolution of rotational ambiguity for three-component systems. *Anal. Chem.*, 83(3):836–841, 2011.

[48] A. Golshan, H. Abdollahi, M. Maeder. The reduction of rotational ambiguity in soft-modeling by introducing hard models. *Anal. Chim. Acta*, 709(0):32–40, 2012.

[49] G. Golub, C. van Loan. *Matrix computations, 4th edition*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2012.

[50] T. Gonçalves, L. Rosa, R. Gonçalves, A. Torquato, P. Março, S. Gomes, M. Matsushita, P. Valderrama.

Monitoring the oxidative stability of monovarietal extra virgin olive oils by UV–Vis spectroscopy and MCR–ALS. *Food Anal. Methods*, S. 1–8, 2018.

[51] D. Gregory, N. Pullman. Semiring rank: Boolean rank and nonnegative rank factorizations. *J. Combin. Inform. System Sci*, 8(3):223–233, 1983.

[52] X. Guan, W. Wang, X. Zhang. Fast intrusion detection based on a non-negative matrix factorization model. *J. Netw. Comput. Appl.*, 32(1):31–44, 2009.

[53] E. Hairer, G. Wanner. Stiff differential equations solved by Radau methods. *J. Comput. Appl. Math.*, 111(1-2):93–111, 1999.

[54] E. Hairer, G. Wanner, S. Nørsett. *Solving ordinary differential equations I, 2nd edition*. Springer, 2002.

[55] J. Handamard. Lectures on the cauchy problems in linear partial differential equations,(1923), 1923.

[56] P. Hansen. The truncated SVD as a method for regularization. *BIT Numer. Math.*, 27(4):534–553, Dec 1987.

[57] P. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

[58] P. Hansen, D. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. and Stat. Comput.*, 14(6):1487–1503, 1993.

[59] M. Hesse, H. Meier, B. Zeeh. *Spektroskopische Methoden in der organischen Chemie*. Georg Thieme Verlag, 2005.

[60] P. Hoyer. Non-negative matrix factorization with sparseness constraints. *J. Mach. Learning Res.*, 5:1457–1469, 2004.

[61] S. Hugelier, O. Devos, C. Ruckebusch. On the implementation of spatial constraints in multivariate curve resolution alternating least squares for hyperspectral image analysis. *J. Chemom.*, 29(10):557–561, 2015.

[62] W. Jackson, J. Harrowfield, P. Vowles. Consecutive, irreversible first-order reactions. Ambiguities and practical aspects of kinetic analyses. *Int. J. Chem. Kinet.*, 9(4):535–548, 1977.

[63] J. Jaumot, A. de Juan, R. Tauler. MCR-ALS GUI 2.0: New features and applications. *Chemom. Intell. Lab. Syst.*, 140(Supplement C):1 – 12, 2015.

[64] J. Jaumot, P. Gemperline, A. Stang. Non-negativity constraints for elimination of multiple solutions in fitting of multivariate kinetic models to spectroscopic data. *J. Chemom.*, 19(2):97–106, 2005.

[65] A. Jürß. Über nichtnegative Matrixfaktorisierungen und geometrische Algorithmen zur Approximation ihrer Lösungsmengen. *Dissertationsschrift, Universität Rostock*, 2017.

[66] H. Kim, H. Park. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix. Anal. Appl.*, 30:713–730, 2008.

[67] P. Kosmol. *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben*. Teubner Studienbücher, 1989.

[68] C. Kubis. Untersuchungen zur Kinetik der Hydroformylierung mit Phosphit-modifizierten Rhodiumkatalysatoren unter Einsatz der in situ IR-Spektroskopie. *Dissertation, Universität Rostock*, 2004.

[69] C. Kubis, W. Baumann, E. Barsch, D. Selent, M. Sawall, R. Ludwig, K. Neymeyr, D. Hess, R. Franke, A. Börner. Investigation into the equilibrium of Iridium catalysts for the hydroformylation of olefins by combining in situ high-pressure FTIR- and NMR-spectroscopy. *ACS Catal.*, 4:2097–2108, 2014.

[70] C. Kubis, M. Sawall, A. Block, K. Neymeyr, R. Ludwig, A. Börner, D. Selent. An operando FTIR spectroscopic and kinetic study of carbon monoxide pressure influence on rhodium-catalyzed olefin hydroformylation. *Chem. Eur. J.*, 20(37):11921–11931, 2014.

[71] C. Kubis, D. Selent, M. Sawall, R. Ludwig, K. Neymeyr, W. Baumann, R. Franke, A. Börner. Exploring between the extremes: Conversion dependent kinetics of phosphite-modified hydroformylation catalysis. *Chem. Eur. J.*, 18(28):8780–8794, 2012.

[72] O. Kvalheim, F. Brakstad, Y. Liang. Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise. *Anal. Chem.*, 66(1):43–51, 1994.

[73] H. Laurberg. Uniqueness of non-negative matrix factorization. In *14th Stat. Signal Processing Workshop*, S. 44–48. IEEE, 2007.

[74] H. Laurberg, M. Christensen, M. Plumbley, L. Hansen, S. Jensen. Theorems on positive data: On the uniqueness of NMF. *Comput. Intell. Neurosci.*, 2008, 2008.

[75] W. Lawton, E. Sylvestre. Self modelling curve resolution. *Technometrics*, 13:617–633, 1971.

[76] D. Lee, H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[77] D. Lee, H. Seung. Algorithms for non-negative matrix factorization. *Adv. Neural. Inf. Process. Syst.*, 13:556–562, 2001.

[78] P. Lerman. Fitting segmented regression models by grid search. *J. Royal Stat. Soc. Series C (Applied Statistics)*, 29(1):77–84, 1980.

[79] W. Liu, J. Yi. Existing and new algorithms for nonnegative matrix factorization. Technical Report, University of Texas, Austin, 2003.

[80] M. Maeder, P. King. Reactlab, 2009.

[81] M. Maeder, Y. Neuhold. *Practical data analysis in chemistry.* Elsevier, Amsterdam, 2007.

[82] E. Malinowski. *Factor analysis in chemistry, 3rd edition.* Wiley, New York, 2002.

[83] H. Mauser, G. Gauglitz. *Photokinetics: theoretical fundamentals and applications*, Bd. 36. Elsevier, 1998.

[84] W. Milligan, D. Mullica, D. Pennington, C. Lok, D. Kwong. Application of nonlinear least squares analysis on three different consecutive irreversible first order kinetic proceses. *Comput. Chem.*, 8(4):285–298, 1984.

[85] H. Minc. *Nonnegative matrices.* John Wiley & Sons, New York, 1988.

[86] K. Neymeyr, M. Sawall. On the set of solutions of the nonnegative matrix factorization problem. *SIAM J. Matrix. Anal. Appl.*, 39(2):1049–1069, 2018.

[87] K. Neymeyr, M. Sawall, D. Hess. Pure component spectral recovery and constrained matrix factorizations: Concepts and applications. *J. Chemom.*, 24:67–74, 2010.

[88] A. Ozerov, C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Trans. Audio Speech Lang. Process.*, 18(3):550–563, 2009.

[89] P. Paatero, U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

[90] M. Papageorgiou, M. Leibold, M. Buss. *Optimierung.* Springer, 1991.

[91] V. Pauca, J. Piper, R. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra Appl.*, 416(1):29–47, 2006.

[92] V. Pauca, F. Shahnaz, M. Berry, R. Plemmons. Text mining using non-negative matrix factorizations. In *Proc. SIAM Int. Conf. Data Min.*, S. 452–456. SIAM, 2004.

[93] O. Perron. Zur Theorie der Matrices. *Math. Ann.*, 64(2):248–263, 1907.

[94] D. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9(1):84–97, 1962.

[95] R. Rajkó, K. István. Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry. *J. Chemom.*, 19(8):448–463, 2005.

[96] C. Ruckebusch, M. Sliwa, P. Pernot, A. de Juan, R. Tauler. Comprehensive data analysis of femtosecond transient absorption spectra: A review. *J. Photochem. Photobiol., C*, 13(1):1–27, 2012.

[97] A. Ruckstuhl, M. Jacobson, R. Field, J. Dodd. Baseline subtraction using robust local regression estimation. *J. Quant. Spectrosc. Radiat. Transfer*, 68(2):179–193, 2001.

[98] E. Sandvol, D. Seber, A. Calvert, M. Barazangi. Grid search modeling of receiver functions: Implications for crustal structure in the middle east and north africa. *J. Geophys. Res.: Solid Earth*, 103(B11):26899–26917, 1998.

[99] A. Savitzky, M. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36(8):1627–1639, 1964.

[100] M. Sawall. Regularisierte nichtnegative Matrixfaktorisierungen und ihre Anwendungen in der Spektroskopie. *Dissertationsschrift, Universität Rostock*, 2011.

[101] M. Sawall. Analyse und Berechnung niedrigdimensionaler Darstellungen von Lösungsmengen zur nichtnegativen Matrixfaktorisierung. *Habilitationsschrift, Universität Rostock*, 2018.

[102] M. Sawall, A. Jürß, H. Schröder, K. Neymeyr. On the analysis and computation of the area of feasible solutions for two-, three-, and four-component systems. In *Data Handling in Science and Technology*, Bd. 30, S. 135–184. Elsevier, 2016.

[103] M. Sawall, A. Jürß, H. Schröder, K. Neymeyr. Simultaneous construction of dual Borgen plots. I: The case of noise-free data. *J. Chemom.*, 31(12):e2954, 2017.

[104] M. Sawall, C. Kubis, E. Barsch, D. Selent, A. Boerner, K. Neymeyr. Peak group analysis for the extraction of pure component spectra. *J. Iran. Chem. Soc.*, 13(2):191–205, 2016.

[105] M. Sawall, C. Kubis, D. Selent, A. Börner, K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: concepts and applications. *J. Chemom.*,

27(5):106–116, 2013.

[106] M. Sawall, A. Moog, C. Kubis, H. Schröder, D. Selent, R. Franke, A. Brächer, A. Börner, K. Neymeyr. Simultaneous construction of dual Borgen plots. II: Algorithmic enhancement for applications to noisy spectral data. *J. Chemom.*, 32(6):e3012, 2018.

[107] M. Sawall, K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. II: Theoretical foundation, inverse polygon inflation, and FAC-PACK implementation. *J. Chemom.*, 28(8):633–644, 2014.

[108] M. Sawall, K. Neymeyr. A ray casting method for the computation of the area of feasible solutions for multicomponent systems: Theory, applications and facpack-implementation. *Anal. Chim. Acta*, 960:40–52, 2017.

[109] M. Sawall, N. Rahimdoust, C. Kubis, H. Schroeder, D. Selent, D. Hess, H. Abdollahi, R. Franke, A. Boerner, K. Neymeyr. Soft constraints for reducing the intrinsic rotational ambiguity of the area of feasible solutions. *Chemom. Intell. Lab. Syst.*, 149:140–150, 2015.

[110] M. Sawall, H. Schröder, D. Meinhardt, K. Neymeyr. *On the ambiguity underlying multivariate curve resolution methods.* Buchkapitel zur späteren Veröffentlichung in Comprehensive Chemometrics, 2nd Edition, Elsevier.

[111] H. Schröder, C. Ruckebusch, O. Devos, R. Métivier, M. Sawall, D. Meinhardt, K. Neymeyr. Analysis of the ambiguity in the determination of quantum yields from spectral data on a photoinduced isomerization. *Chemom. Intell. Lab. Syst.*, 189:88–95, 2019.

[112] H. Schröder, M. Sawall, C. Kubis, D. Selent, D. Hess, R. Franke, A. Börner, K. Neymeyr. On the ambiguity of the reaction rate constants in multivariate curve resolution for reversible first-order reaction systems. *Anal. Chim. Acta*, 927:21–34, 2016.

[113] H. Schröder. Kinetische Modellierung für multivariate Faktormethoden. *Masterthesis, Universität Rostock*, 2013.

[114] L. Shampine, M. Reichelt. The matlab ode suite. *SIAM J. Sci. Comput.*, 18(1):1–22, 1997.

[115] A. Skvortsov. Estimation of rotation ambiguity in multivariate curve resolution with charged particle swarm optimization (cPSO-MCR). *J. Chemom.*, 28(10):727–739, 2014.

[116] A. Smilde, H. Hoefsloot, H. Kiers, S. Bijlsma, H. Boelens. Sufficient conditions for unique solutions within a certain class of curve resolution models. *J. Chemom.*, 15(4):405–411, 2001.

[117] J. Smith, F. Smith, K. Booksh. Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) with raman imaging applied to lunar meteorites. *Appl. Spectrosc.*, 72(3):404–419, 2018.

[118] D. Summers, J. Scott. Systems of first-order chemical reactions. *Math. Comput. Model.*, 10(12):901–909, 1988.

[119] R. Tauler. Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst.*, 30(1):133–146, 1995.

[120] R. Tauler, A. Izquierdo-Ridorsa, E. Casassas. Simultaneous analysis of several spectroscopic titrations with self-modelling curve resolution. *Chemom. Intell. Lab. Syst.*, 18(3):293–300, 1993.

[121] R. Tauler, I. Marqués, E. Casassas. Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid–base titration of salicylic acid at three excitation wavelengths. *J. Chemom.*, 12(1):55–75, 1998.

[122] R. Tauler, A. Smilde, B. Kowalski. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.*, 9(1):31–58, 1995.

[123] L. Thomas. Rank factorization of nonnegative matrices (A. Berman). *SIAM Rev.*, 16(3):393–394, 1974.

[124] A. Tikhonov, V. Arsenin. *Solutions of ill-posed problems*, Bd. 14. Winston Washington, DC, 1977.

[125] S. Vajda, H. Rabitz. Identifiability and distinguishability of first-order reaction systems. *J. Phys. Chem.*, 92(3):701–707, 1988.

[126] S. Vajda, H. Rabitz. Identifiability and distinguishability of general reaction systems. *J. Phys. Chem.*, 98(20):5265–5271, 1994.

[127] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, M. Maeder. On rotational ambiguity in model-free analyses of multivariate data. *J. Chemom.*, 20(6-7):302–310, 2006.

[128] E. Walter. *Identifiability of parametric models.* Elsevier, 2014.

[129] S. Wang, C. Deng, W. Lin, G. Huang, B. Zhao. NMF-based image quality assessment using extreme learning machine. *IEEE Trans. Cybern.*, 47(1):232–243, 2016.

[130] G. Wanner, E. Hairer. Solving ordinary differential equations II. *Stiff and differential-algebraic problems*, 1991.

[131] X. Zhang, R. Tauler. Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. *Anal. Chim. Acta*, 762:25–38, 2013.

[132] C. Zheng, T. Ng, L. Zhang, C. Shiu, H. Wang. Tumor classification based on non-negative matrix factorization using gene expression data. *IEEE Trans. Nanobiosci.*, 10(2):86–93, 2011.

A. Anhang

A.1. Schwarm-Algorithmus

Im Abschnitt 6.1.1 wurde der Grid Search Algorithmus vorgestellt. Durch die Wahl eines festen Gitters können Niveaumengen $\mathcal{N}$, siehe (6.1), häufig nicht hinreichend genau approximiert werden. Die Verwendung verhältnismäßig großer Fehlertoleranzen ist nötig. Die Idee des nun vorgestellte Schwarm-Algorithmus basiert auf dem Grid Search Algorithmus, jedoch werden die Gitterpunkte nicht nur ausgewertet, sondern sie sind Startiterierten einer anschließenden Optimierung [33]. Hierdurch kann der Fehler der Approximation deutlich gesenkt werden. Darüber hinaus wird eine möglichst gleichmäßige Verteilung der finalen Iterierten angestrebt.

Hierzu wird zunächst eine Modifikation $f(k)$, siehe (6.2), durchgeführt:

$$\tilde{f}(k, d, G_0) := f(k) + \sum_{k_0 \in G_0} \min(\|k - k_0\|_2 - d, 0).$$

Dabei entspricht d dem angestrebten Mindestabstand zwischen jeweils zwei finalen Iterierten. Die Menge G_0 ist zunächst leer. Nach jeder Minimierung geprüft, ob für die finale Iterierte k^* der Funktionswert $\tilde{f}(k^*, d, G_0)$ kleiner als eine vorgegebene Fehlertoleranz ε ist. In diesem Fall wird k^* der Menge G_0 hinzugefügt. Die Vorgehensweise ist in Algorithmus 3 zusammengefasst.

Algorithmus 3 Schwarm-Algorithmus

Input: Gitterpunkte g_i mit $G_d = \{g_i \in \mathbb{R}^q : i = 1, \ldots, n_G^q\}$, Mindestabstand $d \geq 0$, Fehlertoleranz $\varepsilon \geq 0$
Output: Approximation von $\mathcal{N}$ durch G_o

$G_o = \emptyset$
for all $k \in G_d$ **do**
  Bestimme k^* durch Minimierung von $\tilde{f}(k, d, G_0)$ mit Startvektor k
  if $\tilde{f}(k^*, d, G_0) \leq \varepsilon$ **then**
    $G_0 = G_0 \cup k^*$
  end if
end for

A.2. Anpassung eines kinetischen Modells zur Analyse einer Rhodium-katalysierte Hydroformylierung

Spektroskopischer Datensatz 4 (Rhodium katalysierte Hydroformylierung [71])**.** Es handelt sich um eine FTIR-Messserie mit $n = 1353$ Spektren zu jeweils $m = 610$

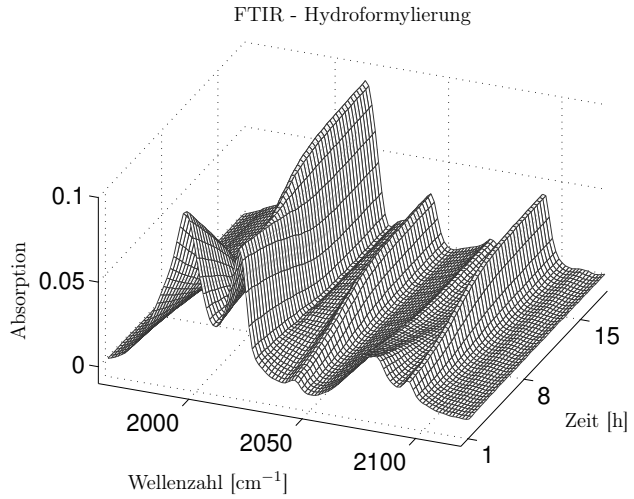


Abbildung A.1.: Darstellung der Spektrenfolge D des spektroskopischen Datensatzes 4.

Wellenzahlen. Im zugrunde liegenden Frequenzfenster $[1960\text{cm}^{-1}, 2120\text{cm}^{-1}]$ absorbieren $s = 3$ Komponenten. Diese sind der Reaktant ($\mathcal{R}$) *Olefin*, ein *Hydridokomplex* als Katalysator($\mathcal{X}$) und ein *Acylkomplex* als Intermediat($\mathcal{Y}$). Der Konzentrationsverlauf des Produkts($\mathcal{P}$), einem *Aldehyd*, konnte aufgrund einer isolierten Bande bei 1694cm^{-1} vorab bestimmt werden und liegt als Vektor $C_{\mathcal{P}} \in \mathbb{R}^n$ vor. Der Vektor der Anfangskonzentrationen lautet

$$c_0 := (c_{\mathcal{R}}(0), c_{\mathcal{X}}(0), c_{\mathcal{Y}}(0), c_{\mathcal{P}}(0))^T = (8.9265 \cdot 10^{-1}, 3.0113 \cdot 10^{-4}, 0, 0)^T$$

in mol. Eine Darstellung der Matrix D ist in der Abbildung A.1 zu sehen.

Für den spektroskopischen Datensatz 4 kann die Michaelis-Menten Kinetik

$$\mathcal{R} + \mathcal{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{Y} \xrightarrow{k_3} \mathcal{X} + \mathcal{P}$$

angenommen werden. Da vier Komponenten durch die Kinetik beschrieben werden, aber nur drei davon im Frequenzfenster der Messung absorbieren, handelt es sich um ein Daten-defizitäres Problem. Die Optimierung basiert auf der Zielfunktion $F_{\text{hard},s<z}(k)$ aus Abschnitt 3.2. Zusätzlich wird der bekannte Konzentrationsverlauf $C_{\mathcal{P}}$ in die Analyse einbezogen und auf Konsistenz mit der Kinetik geprüft, sodass die additiv zusammengesetzte Zielfunktion

$$F_{\text{opt}}(k) = F_{\text{hard},s<z}(k) + \delta_2 \sum_{i=1}^{m} \left(\frac{(C_{\mathcal{P}})_i - (C^{\text{dgl}}(k))_{i,4}}{\max_l (C_{\mathcal{P}})_l} \right)^2$$

lautet. Die Optimierung wurde mit $k^{(0)} = (40\text{mol}^{-1}\text{min}^{-1}, 15\text{min}^{-1}, 15\text{min}^{-1})^T$ in s^{-1} initialisiert und $\gamma_1 = \gamma_2 = \gamma_3 = \delta_1 = \delta_2 = 1$ gewählt. Für die optimierten Parameter $k^* = (62.29\text{mol}^{-1}\text{min}^{-1}, 3.90\text{min}^{-1}, 9.58\text{min}^{-1})^T$ lautet der Zielfunktionswert $F_{\text{opt}}(k^*) = 0.306$ und die Ergebnisse sind in Abbildung A.2 dargestellt. Der relativ große Wert von $F_{\text{opt}}(k^*)$ ergibt sich aus der Nichtnegativität-Forderung an der Faktor S. In der Abbildung sind die starken negativen Anteile deutlich zu erkennen. Sie müssen toleriert werden

und resultieren aus einer notwendigen Datenvorbehandlung, welche die Subtraktion eines Lösungsmittelspektrums beinhaltet.

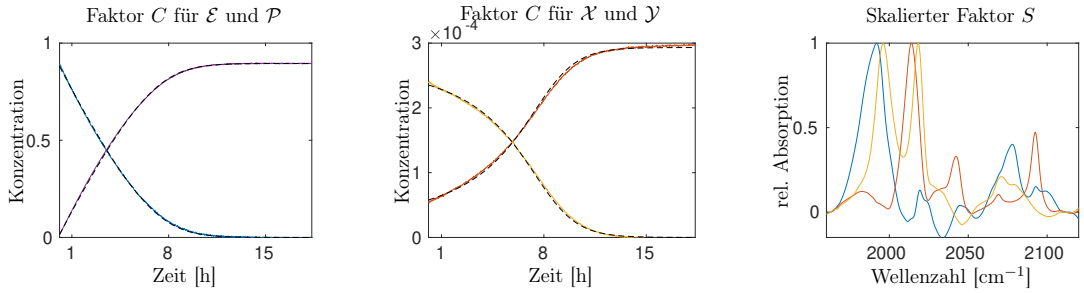


Abbildung A.2.: Darstellung der Faktoren C^* und S^* der optimierten Parameter k^* zu Datensatz 4. Links und in der Mitte sind die Konzentrationsverläufe zu sehen, welche aufgrund des Unterschieds in der Größenordnung aufgeteilt wurden. In lila ist der bereits bekannte Verlauf des Produkts $\mathcal{P}$ dargestellt. Die gestrichelten Linien zeigen die Auswertung des kinetischen Modells zu den optimierten Parametern $C^{\mathrm{dgl}}(k^*)$. Rechts ist der Faktor S^* dargestellt. Zur qualitativen Veranschaulichung wurden die Maxima der einzelnen Spektren auf 1 skaliert.

A.3. Ergänzungen zur Menge $\tilde{\mathcal{K}}^{+}_{0.014,0.022}$ aus Abbildung 7.10

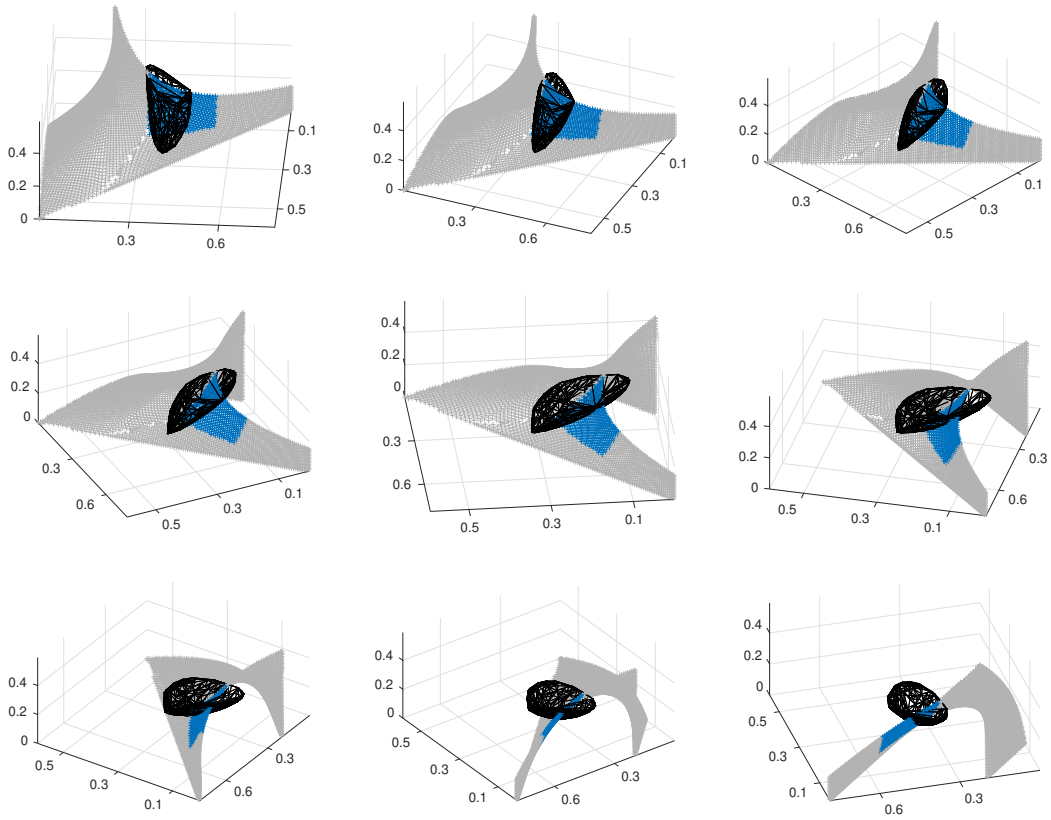


Abbildung A.3.: Für eine bessere räumliche Vorstellung der Menge $\tilde{\mathcal{K}}^{+}_{0.014,0.022}$ wurde die entsprechende Grafik aus Abbildung 7.10 in 20-Grad-Schritten rotiert. Die Anordnung erfolgt zeilenweise und jeweils von links nach rechts.

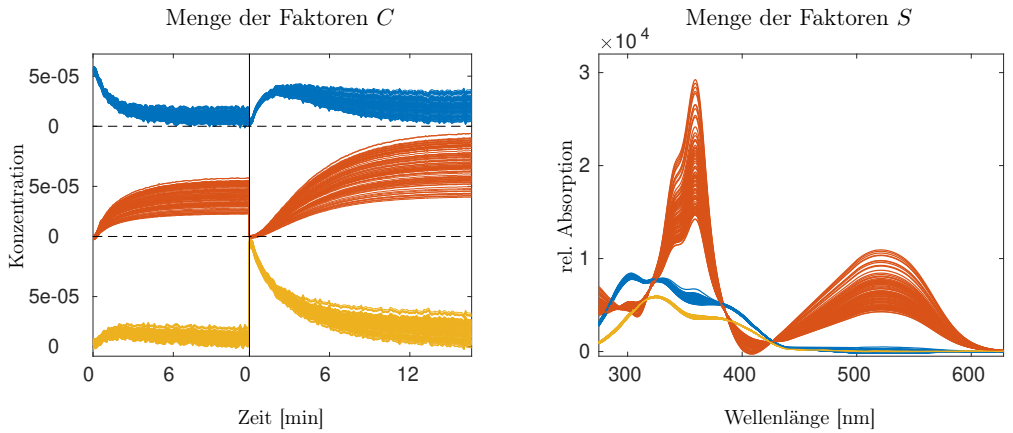


Abbildung A.4.: Darstellung der Menge von Faktoren C (links) und S (rechts) zu 2% zufällig ausgewählten Elementen der Approximation $\tilde{\mathcal{K}}^{+}_{0.014,0.022}$ aus Abbildung 7.10.

Selbstständigkeitserklärung

Ich versichere eidesstattlich durch eigenhändige Unterschrift, dass ich die Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, habe ich als solche kenntlich gemacht.

Rostock, ...

(Henning Schröder)